

00/10/00

03-02-00

A

Customer No. 20350

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
(415) 576-0200

ASSISTANT COMMISSIONER FOR PATENTS
BOX PATENT APPLICATION
Washington, D.C. 20231

Sir:

Transmitted herewith for filing under 37 CFR 1.53(b) is the

- ☐ patent application of
☐ continuation patent application of
☐ divisional patent application of
☒ continuation-in-part patent application of 09/193,931

Attorney Docket No. 023070-077630US

Client Ref No. 99-177-5

"Express Mail" Label No. EL378166615US

Date of Deposit: March 1, 2000

I hereby certify that this is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above, addressed to:

Assistant Commissioner for Patents
Washington, D.C. 20231

By:

Inventor(s)/Applicant Identifier: John Harada, Tamar Lotan, Masa-aki Ohto, Robert B. Goldberg, Robert L. Fischer, Anhthu Bui, Raymond Kwong

For: LEAFY COTYLEDON1 GENES AND THEIR USES

☒ This application claims priority from each of the following Application Nos./filing dates:

The present application is a Continuation-In-Part ("CIP") of U.S. Patent Application No. 09/193,931, filed November 17, 1998, which is a CIP of U.S. Patent Application No. 09/103,478, filed June 24, 1998, which is a CIP of U.S. Patent Application No. 09/026,221, filed February 19, 1998, which is a CIP of U.S. Patent Application 08/804,534, filed February 21, 1997.

the disclosure(s) of which is (are) incorporated by reference.

Enclosed are:

- ☒ 43 page(s) of specification
☒ 5 page(s) of claims
☒ 1 page of Abstract
☒ 1 sheet(s) of ☐ formal ☒ informal drawing(s).
☒ 21 page(s) of sequence listing
☐ A ☐ signed ☐ unsigned Declaration & Power of Attorney
☒ A ☐ signed ☒ unsigned Declaration.
☐ A Power of Attorney by Assignee with Certificate Under 37 CFR Section 3.73(b).
☐ A verified statement to establish small entity status under 37 CFR 1.9 and 37 CFR 1.27 ☐ is enclosed ☐ was filed in the prior application and small entity status is still proper and desired.

| | (Col. 1) | (Col. 2) |
|---|-----------|-----------|
| FOR: | NO. FILED | NO. EXTRA |
| BASIC FEE | | |
| TOTAL CLAIMS | 69 - 20 | = *49 |
| INDEP. CLAIMS | 12 - 3 | = *9 |
| <input type="checkbox"/> MULTIPLE DEPENDENT CLAIM PRESENTED | | |

* If the difference in Col. 1 is less than 0, enter "0" in Col. 2.

SMALL ENTITY

| RATE | FEE |
|--------------|----------|
| | \$345.00 |
| x \$9.00 = | |
| x \$39.00 = | |
| + \$130.00 = | |
| TOTAL | |

OTHER THAN SMALL ENTITY

| RATE | FEE |
|--------------|------------|
| | \$690.00 |
| x \$18.00 = | \$882.00 |
| x \$78.00 = | \$702.00 |
| + \$260.00 = | |
| TOTAL | \$2,274.00 |

Please charge Deposit Account No. 20-1430 as follows:

- ☒ Filing fee \$2,274.00
☒ Any additional fees associated with this paper or during the pendency of this application.
☐ The issue fee set in 37 CFR 1.18 at or before mailing of the Notice of Allowance, pursuant to 37 CFR 1.311(b)

- ☐ A check for \$_____ is enclosed.
 2 extra copies of this sheet are enclosed.

Telephone:
(415) 576-0200

Facsimile:
(415) 576-0300

Respectfully submitted,
TOWNSEND and TOWNSEND and CREW LLP

Kevin Bastian
Reg No.: 34,774
Attorneys for Applicant

PATENT APPLICATION

LEAFY COTYLEDON1 GENES AND THEIR USES

Inventors: John Harada, a citizen of the United States
residing at: 115 Grande Avenue, Davis, CA 95616

Tamar Lotan, a citizen of Israel
residing at: 408 Russell Park, #4, Davis, CA 95616

Masa-aki Ohto, a citizen of Japan
residing at: 2-5-1 Tatsumi-minami, #1-23, Okazaki, Japan 444

Robert B. Goldberg, a citizen of the United States
residing at: 21907 Canon Drive, Topanga, CA 90290

Robert L. Fischer, a citizen of the United States
residing at: 1423 Scott Street, El Cerrito, CA 94530

Anhthu Bui, a citizen of the United States
residing at: 6 National Place, Irvine, CA 92602-0703

Raymond Kwong, a citizen of the United States
residing at: 1915 26th Street, Sacramento, CA 95816-7303

Assignee:

Regents of the University of California

TOWNSEND and TOWNSEND and CREW
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
(415) 576-0200

LEAFY COTYLEDON1 GENES AND THEIR USES

CROSS REFERENCE TO RELATED APPLICATIONS

The present application is a Continuation-In-Part ("CIP") of United States Patent Application Serial Number (USSN) 09/193,931, filed November 17, 1998, which is a CIP of USSN 09/103,478, filed June 24, 1998, which is a CIP of USSN 09/026,221, filed February 19, 1998, which is a CIP of USSN 08/804,534, filed February 21, 1997. Each of the aforementioned applications is explicitly incorporated herein by reference in their entirety and for all purposes.

FIELD OF THE INVENTION

The present invention is directed to plant genetic engineering. In particular, it relates to new embryo-specific genes useful in improving agronomically important plants.

BACKGROUND OF THE INVENTION

Embryogenesis in higher plants is a critical stage of the plant life cycle in which the primary organs are established. Embryo development can be separated into two main phases: the early phase in which the primary body organization of the embryo is laid down and the late phase which involves maturation, desiccation and dormancy. In the early phase, the symmetry of the embryo changes from radial to bilateral, giving rise to a hypocotyl with a shoot meristem surrounded by the two cotyledonary primordia at the apical pole and a root meristem at the basal pole. In the late phase, during maturation the embryo achieves its maximum size and the seed accumulates storage proteins and lipids. Maturation is ended by the desiccation stage in which the seed water content decreases rapidly and the embryo passes into metabolic quiescent state. Dormancy ends with seed germination, and development continues from the shoot and the root meristem regions.

The precise regulatory mechanisms which control cell and organ differentiation during the initial phase of embryogenesis are largely unknown. The plant hormone abscisic acid (ABA) is thought to play a role during late embryogenesis, mainly in the maturation stage by inhibiting germination during embryogenesis (Black, M. (1991). In *Abscissic Acid: Physiology and Biochemistry*, W. J. Davies and H. G. Jones, eds. (Oxford: Bios Scientific Publishers Ltd.), pp. 99-124) Koornneef, M., and Karssen, C. M. (1994). In *Arabidopsis*, E. M. Meyerowitz and C. R. Somerville, eds. (Cold Spring Harbor: Cold Spring Harbor Laboratory Press), pp. 313-334). Mutations which effect seed development

and are ABA insensitive have been identified in Arabidopsis and maize. The ABA insensitive (*abi3*) mutant of Arabidopsis and the viviparous1 (*vp1*) mutant of maize are detected mainly during late embryogenesis (McCarty, et al., (1989) *Plant Cell* 1, 523-532 and Parcy et al., (1994) *Plant Cell* 6, 1567-1582). Both the VP1 gene and the ABI3 genes have
 5 been isolated and were found to share conserved regions (Giraudat, J. (1995) *Current Opinion in Cell Biology* 7:232-238 and McCarty, D. R. (1995). *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 46:71-93). The VP1 gene has been shown to function as a transcription activator (McCarty, et al., (1991) *Cell* 66:895-906). It has been suggested that ABI3 has a similar function.

10 Another class of embryo defective mutants involves three genes: LEAFY COTYLEDON1 and 2 (*LEC1*, *LEC2*) and FUSCA3 (*FUS3*). These genes are thought to play a central role in late embryogenesis (Baumlein, et al. (1994) *Plant J.* 6:379-387; Meinke, D. W. (1992) *Science* 258:1647-1650; Meinke et al., *Plant Cell* 6:1049-1064; West et al., (1994) *Plant Cell* 6:1731-1745). Like the *abi3* mutant, leafy cotyledon-type mutants are
 15 defective in late embryogenesis. In these mutants, seed morphology is altered, the shoot meristem is activated early, storage proteins are lacking and developing cotyledons accumulate anthocyanin. As with *abi3* mutants, they are desiccation intolerant and therefore die during late embryogenesis. Nevertheless, the immature mutants embryos can be rescued to give rise to mature and fertile plants. However, unlike *abi3* when the immature mutants
 20 germinate they exhibit trichomes on the adaxial surface of the cotyledon. Trichomes are normally present only on leaves, stems and sepals, not cotyledons. Therefore, it is thought that the leafy cotyledon type genes have a role in specifying cotyledon identity during embryo development.

Among the above mutants, the *lec1* mutant exhibits the most extreme
 25 phenotype during embryogenesis. For example, the maturation and postgermination programs are active simultaneously in the *lec1* mutant (West et al., 1994), suggesting a critical role for LEC1 in gene regulation during late embryogenesis.

In spite of the recent progress in defining the genetic control of embryo
 30 development, further progress is required in the identification and analysis of genes expressed specifically in the embryo and seed. Characterization of such genes would allow for the genetic engineering plants with a variety of desirable traits. For instance, modulation of the expression of genes which control embryo development may be used to alter traits such as accumulation of storage proteins in leaves and cotyledons. Alternatively, promoters from

embryo or seed-specific genes can be used to direct expression of desirable heterologous genes to the embryo or seed. The present invention addresses these and other needs.

SUMMARY OF THE INVENTION

5 The present invention is based, in part, on the isolation and characterization of LEC1 genes. The invention provides isolated nucleic acid molecules comprising a LEC1 polynucleotide sequence which is at least 68% identical to the B domain of SEQ ID NO:2.

The invention also provides expression cassettes comprising a promoter operably linked to a heterologous polynucleotide sequence or complement thereof, encoding
10 a LEC1 polypeptide comprising a sequence which is at least 68% identical to the B domain of SEQ ID NO:2. In some embodiments, the polynucleotide sequence is heterologous to any element in the expression cassette. In a preferred embodiment, the B domain comprises a polypeptide between about amino acid residue 28 and amino acid residue 117 of SEQ ID NO:2. In a more preferred embodiment, the B domain comprises a polypeptide sequence
15 with an amino terminus at amino acid residues 28-35 and a carboxy terminus at amino acid residues 103-117 of SEQ ID NO:2.

In particularly preferred embodiments, the LEC1 polypeptide is shown in SEQ ID NO:20 or 22. Such LEC1 polypeptides can be encoded by the polynucleotide sequences shown in SEQ ID NO:19 or SEQ ID NO:21, respectively. In another embodiment, the LEC1
20 polypeptide is a fusion between two or more LEC1 polypeptides of polypeptide subsequences.

The expression cassette comprises a promoter operably linked to the LEC1 polynucleotide or its complement. For example, the promoter can be a constitutive promoter. Alternatively, the promoter can be a promoter from a LEC1 gene. For instance, the LEC1
25 promoter can be from about nucleotide 1 to about nucleotide 1998 of SEQ ID NO:3. In one embodiment, the heterologous polynucleotide can be linked to the promoter in the antisense orientation. In another embodiment, the promoter is SEQ ID NO:23. The promoter can further comprise SEQ ID NO:24.

In another embodiment, the invention provides an expression cassette
30 comprising a promoter operably linked to a heterologous polynucleotide sequence, or complement thereof, encoding a LEC1 polypeptide comprising a subsequence at least 90% identical to the A or C domain of a LEC1 polypeptide. The polynucleotide sequence can be heterologous to any element in the expression cassette. Such expression cassettes can encode fusions of two or more LEC1 polypeptides or polypeptide subsequences.

The invention also provides for an expression cassette for the expression of heterologous polypeptides in a plant. The expression cassette comprises a LEC1 promoter operably linked to a heterologous polynucleotide. In some embodiments, the LEC1 promoter is at least 70% identical to SEQ ID NO:23. In some embodiments, the expression cassette promoter comprises a promoter at least 70% identical to SEQ ID NO:24. Preferably, the promoter comprises the sequence displayed in SEQ ID NO:24.

The invention also provides an isolated nucleic acid or complement thereof, encoding a LEC1 polypeptide comprising a subsequence at least 68% identical to the B domain of SEQ ID NO:2, with the proviso that the nucleic acid is not clone MNJ7. In a preferred embodiment, the B domain comprises a polypeptide sequence with an amino terminus at amino acids 28-35 and a carboxy terminus at amino acids 103-117 of SEQ ID NO:2. In another embodiment, the LEC1 polypeptide is shown in SEQ ID NO: 20 or SEQ ID NO:22. Such LEC1 polypeptides can be encoded by the polynucleotide sequences shown in SEQ ID NO:19 or SEQ ID NO:21, respectively. In another embodiment, the LEC1 polypeptide is a fusion between two or more LEC1 polypeptides of polypeptide subsequences.

The isolated nucleic acid can further comprise a promoter operably linked to the LEC1-encoding nucleic acid. The promoter can be a constitutive promoter. Alternatively, the promoter can be a promoter from a LEC1 gene. For instance, the LEC1 promoter can be from about nucleotide 1 to about nucleotide 1998 of SEQ ID NO:3. In one embodiment, the heterologous polynucleotide can be linked to the promoter in the antisense orientation.

The invention provides a host cell comprising expression cassettes or nucleic acids of the invention. Thus, in one embodiment, the host cells of the invention comprise an expression cassette comprising a promoter operably linked to a heterologous a polynucleotide sequence, or complement thereof, encoding a LEC1 polypeptide with a subsequence at least 68% identical to the B domain of SEQ ID NO:2. In other embodiments, the host cell of the invention comprises an expression cassette comprising a promoter operably linked to a heterologous polynucleotide sequence or complement thereof, encoding a LEC1 polypeptide with a subsequence at least 90% identical to the A or C domain of a LEC1 polypeptide. Other embodiments include hosts cells comprising an expression cassette comprising a promoter at least 70% identical to SEQ ID NO:23 or an isolated nucleic acid comprising a subsequence at least 68% identical to the B domain of SEQ ID NO:2, so long as the nucleic acid is not clone MNJ7.

The invention also provides isolated polypeptides comprising amino acid sequences at least 68% identical to the B domain of SEQ ID NO:2 and capable of exhibiting at least one of the biological activities of the polypeptides encoded in SEQ ID NO:1, SEQ ID NO:19 or SEQ ID NO:21, or a fragment thereof. Antibodies capable of binding the above-described polypeptide are also provided.

Also provided are methods of introducing an isolated nucleic acid into a host cell. The method comprises providing an expression cassette of nucleic acid of the invention as described herein and contacting the expression cassette or nucleic acid with the host cell under conditions that permit insertion of the nucleic acid into the host cell.

The invention also provides transgenic plant cells or plants comprising an expression cassette comprising a promoter operably linked to a heterologous polynucleotide sequence, or complement thereof, encoding a LEC1 polypeptide comprising a sequence which is at least 68% identical to the B domain of SEQ ID NO:2. In a preferred embodiment, the LEC1 polypeptide is shown in SEQ ID NO: 20 or SEQ ID NO:22. Such LEC1 polypeptides can be encoded by the polynucleotide sequences shown in SEQ ID NO:19 or SEQ ID NO:21, respectively. The invention also provides plants that are regenerated from the plant cells discussed above.

The expression cassette promoter can be a constitutive promoter. Alternatively, the promoter can be a promoter from a LEC1 gene. For instance, the LEC1 promoter can be from about nucleotide 1 to about nucleotide 1998 of SEQ ID NO:3. In one embodiment, the heterologous polynucleotide can be linked to the promoter in the antisense orientation. In another embodiment, the promoter is SEQ ID NO:23. The promoter can also further comprise SEQ ID NO:24.

The invention also provides an expression cassette for the expression of a heterologous polynucleotide in a plant cell, comprising a promoter polynucleotide at least 70% identical to SEQ ID NO:23, wherein the promoter polynucleotide is operably linked to a heterologous polynucleotide. In one embodiment, the promoter polynucleotide is SEQ ID NO:23. The promoter can also further comprise a polynucleotide at least 70% identical to SEQ ID NO:24. In a preferred embodiment, the promoter comprises SEQ ID NO:24.

The invention also provides methods of modulating transcription comprising, introducing into the plant an expression cassette containing a plant promoter operably linked to a heterologous *LEC1* polynucleotide, the heterologous *LEC1* polynucleotide encoding a LEC1 polypeptide comprising a subsequence at least 68% identical to the B domain of SEQ ID NO:2 and detecting a plant with modulated transcription. Embodiments of these methods

include where the LEC1 polynucleotide is SEQ ID NO:2, SEQ ID NO:20 or SEQ ID NO:22. In other embodiments, the LEC1 polypeptides are encoded by SEQ ID NO:1, SEQ ID NO:19 or SEQ ID NO:21. Preferred embodiments of the invention include the method where transcription modulation results in induction of embryonic characteristics in a plant. In an alternative embodiment, transcription modulation results in induction of seed development.

The invention also provides a method of detecting a nucleic acid in a sample. The method comprises providing an isolated LEC1 nucleic acid molecule comprising a polynucleotide sequence, or complement thereof, encoding a LEC1 polypeptide with a subsequence at least 68% identical to the B domain of SEQ ID NO:2., contacting the isolated nucleic acid molecule with a sample under conditions which permit a comparison of the sequence of the isolated nucleic acid molecule with the sequence of DNA in the sample; and analyzing the result of the comparison. In some embodiments, the isolated nucleic acid molecule and the sample are contacted under conditions which permit the formation of a duplex between complementary nucleic acid sequences.

Definitions

The phrase "nucleic acid" refers to a single or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. Nucleic acids may also include modified nucleotides that permit correct read through by a polymerase and do not alter expression of a polypeptide encoded by that nucleic acid.

The phrase "polynucleotide sequence" or "nucleic acid sequence" includes both the sense and antisense strands of a nucleic acid as either individual single strands or in the duplex. It includes, but is not limited to, self-replicating plasmids, chromosomal sequences, and infectious polymers of DNA or RNA.

The phrase "nucleic acid sequence encoding" refers to a nucleic acid which directs the expression of a specific protein or peptide. The nucleic acid sequences include both the DNA strand sequence that is transcribed into RNA and the RNA sequence that is translated into protein. The nucleic acid sequences include both the full length nucleic acid sequences as well as non-full length sequences derived from the full length sequences. It should be further understood that the sequence includes the degenerate codons of the native sequence or sequences which may be introduced to provide codon preference in a specific host cell.

The term "promoter" refers to a region or sequence determinants located upstream or downstream from the start of transcription and which are involved in recognition and binding of RNA polymerase and other proteins to initiate transcription. A "plant

promoter" is a promoter capable of initiating transcription in plant cells. Such promoters need not be of plant origin, for example, promoters derived from plant viruses, such as the CaMV35S promoter, can be used in the present invention.

The term "plant" includes whole plants, shoot vegetative organs/structures (e.g. leaves, stems and tubers), roots, flowers and floral organs/structures (e.g. bracts, sepals, petals, stamens, carpels, anthers and ovules), seed (including embryo, endosperm, and seed coat) and fruit (the mature ovary), plant tissue (e.g. vascular tissue, ground tissue, and the like) and cells (e.g. guard cells, egg cells, trichomes and the like), and progeny of same. The class of plants that can be used in the method of the invention is generally as broad as the class of higher and lower plants amenable to transformation techniques, including angiosperms (monocotyledonous and dicotyledonous plants), gymnosperms, ferns, and multicellular algae. It includes plants of a variety of ploidy levels, including aneuploid, polyploid, diploid, haploid and hemizygous.

A polynucleotide sequence is "heterologous to" an organism or a second polynucleotide sequence if it originates from a foreign species, or, if from the same species, is modified from its original form. For example, a promoter operably linked to a heterologous coding sequence refers to a coding sequence from a species different from that from which the promoter was derived, or, if from the same species, a coding sequence which is different from any naturally occurring allelic variants. As defined here, a modified LEC1 coding sequence which is heterologous to an operably linked LEC1 promoter does not include the T-DNA insertional mutants as described in West et al., *The Plant Cell* 6:1731-1745 (1994).

A polynucleotide "exogenous to" an individual plant is a polynucleotide which is introduced into the plant by any means other than by a sexual cross. Examples of means by which this can be accomplished are described below, and include *Agrobacterium*-mediated transformation, biolistic methods, electroporation, in planta techniques, and the like. Such a plant containing the exogenous nucleic acid is referred to here as an R₁ generation transgenic plant. Transgenic plants which arise from sexual cross or by selfing are descendants of such a plant.

As used herein an "embryo-specific gene" or "seed specific gene" is a gene that is preferentially expressed during embryo development in a plant. For purposes of this disclosure, embryo development begins with the first cell divisions in the zygote and continues through the late phase of embryo development (characterized by maturation, desiccation, dormancy), and ends with the production of a mature and desiccated seed. Embryo-specific genes can be further classified as "early phase-specific" and "late phase-

specific". Early phase-specific genes are those expressed in embryos up to the end of embryo morphogenesis. Late phase-specific genes are those expressed from maturation through to production of a mature and desiccated seed.

A “LEC1 polynucleotide” is a nucleic acid sequence comprising (or consisting of) a coding region of about 100 to about 900 nucleotides, sometimes from about 300 to about 630 nucleotides, which hybridizes to SEQ ID NO:1 under stringent conditions (as defined below), or which encodes a LEC1 polypeptide. LEC1 polynucleotides can also be identified by their ability to hybridize under low stringency conditions (e.g., $T_m \sim 40^\circ\text{C}$) to nucleic acid probes having a sequence from position 1 to 81 in SEQ ID NO:1 or from position 355 to 627 in SEQ ID NO:1.

A “promoter from a LEC1 gene” or “LEC1 promoter” will typically be about 500 to about 2000 nucleotides in length, usually from about 750 to 1500. Exemplary promoter sequences are shown as nucleotides 1-1998 of SEQ ID NO:3 or as SEQ ID NO:23. A LEC1 promoter can also be identified by its ability to direct expression in all, or essentially all, proglabular embryonic cells, as well as cotyledons and axes of a late embryo.

A "LEC1 polypeptide" is a sequence of about 50 to about 210, sometimes 100 to 150, amino acid residues encoded by a LEC1 polynucleotide. A full length LEC1 polypeptide and fragments containing a CCAAT binding factor (CBF) domain can act as a subunit of a protein capable of acting as a transcription factor in plant cells. LEC1 polypeptides are often distinguished by the presence of a sequence which is required for binding the nucleotide sequence: CCAAT. In particular, a short region of seven residues (MPIANVI) at residues 34-40 of SEQ ID NO: 3 shows a high degree of similarity to a region that has been shown to required for binding the CCAAT box. Similarly, residues 61-72 of SEQ ID NO: 3 (IQECVSEYISFV) is nearly identical to a region that contains a subunit interaction domain (Xing, et al., (1993) *EMBO J.* 12:4647-4655).

As used herein, a homolog of a particular embryo-specific gene (e.g., SEQ ID NO:1) is a second gene in the same plant type or in a different plant type, which has a polynucleotide sequence of at least 50 contiguous nucleotides which are substantially identical (determined as described below) to a sequence in the first gene. It is believed that, in general, homologs share a common evolutionary past.

"Increased or enhanced LEC1 activity or expression of the *LEC1* gene" refers to an augmented change in LEC1 activity. Examples of such increased activity or expression include the following. LEC1 activity or expression of the *LEC1* gene is increased above the level of that in wild-type, non-transgenic control plants (i.e. the quantity of LEC1 activity or

expression of the *LEC1* gene is increased). *LEC1* activity or expression of the *LEC1* gene is in an organ, tissue or cell where it is not normally detected in wild-type, non-transgenic control plants (i.e. spatial distribution of *LEC1* activity or expression of the *LEC1* gene is increased). *LEC1* activity or expression is increased when *LEC1* activity or expression of the

5 *LEC1* gene is present in an organ, tissue or cell for a longer period than in a wild-type, non-transgenic controls (i.e. duration of *LEC1* activity or expression of the *LEC1* gene is increased).

A "polynucleotide sequence from" a particular embryo-specific gene is a subsequence or full length polynucleotide sequence of an embryo-specific gene which, when

10 present in a transgenic plant, has the desired effect, for example, inhibiting expression of the endogenous gene driving expression of an heterologous polynucleotide. A full length sequence of a particular gene disclosed here may contain about 95%, usually at least about 98% of an entire sequence shown in the Sequence Listing, below.

The term "reproductive tissues" as used herein includes fruit, ovules, seeds,

15 pollen, pistils, flowers, or any embryonic tissue.

In the case of both expression of transgenes and inhibition of endogenous genes (e.g., by antisense, or sense suppression) one of skill will recognize that the inserted polynucleotide sequence need not be identical and may be "substantially identical" to a sequence of the gene from which it was derived. As explained below, these variants are

20 specifically covered by this term.

In the case where the inserted polynucleotide sequence is transcribed and translated to produce a functional polypeptide, one of skill will recognize that because of codon degeneracy a number of polynucleotide sequences will encode the same polypeptide. These variants are specifically covered by the term "polynucleotide sequence from" a

25 particular embryo-specific gene, such as *LEC1*. In addition, the term specifically includes sequences (e.g., full length sequences) substantially identical (determined as described below) with a *LEC1* gene sequence and that encode proteins that retain the function of a *LEC1* polypeptide.

In the case of polynucleotides used to inhibit expression of an endogenous

30 gene, the introduced sequence need not be perfectly identical to a sequence of the target endogenous gene. The introduced polynucleotide sequence will typically be at least substantially identical (as determined below) to the target endogenous sequence.

Two nucleic acid sequences or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the

00645053 030400

same when aligned for maximum correspondence as described below. The term "complementary to" is used herein to mean that the sequence is complementary to all or a portion of a reference polynucleotide sequence.

Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman *Add. APL. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman and Wunsch *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson and Lipman *Proc. Natl. Acad. Sci. (U.S.A.)* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection.

"Percentage of sequence identity" is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 25% sequence identity. Alternatively, percent identity can be any integer from 25% to 100%. More preferred embodiments include at least: 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99%. compared to a reference sequence using the programs described herein; preferably BLAST using standard parameters, as described below. Accordingly, LEC1 sequences of the invention include nucleic acid sequences that have substantial identity to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:19 and SEQ ID NO:21. LEC1 sequences of the invention include polypeptide sequences having substantial identity to SEQ ID NO:2, SEQ ID NO:20 or SEQ ID NO:22. One of skill will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning and the like. Substantial identity of amino acid sequences for these purposes normally means sequence identity of at least 40%. Preferred percent identity of polypeptides can be any integer from 40% to 100%. More preferred embodiments include at

least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99%. Most preferred embodiments include 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74% and 75%. Polypeptides which are "substantially similar" share sequences as noted above except that residue positions which are not identical may differ by conservative amino acid changes. Conservative amino acid substitutions refer to the interchangeability of residues having similar side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulfur-containing side chains is cysteine and methionine. Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, aspartic acid-glutamic acid, and asparagine-glutamine.

Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other, or a third nucleic acid, under stringent conditions. Stringent conditions are sequence dependent and will be different in different circumstances. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically, stringent conditions will be those in which the salt concentration is about 0.02 molar at pH 7 and the temperature is at least about 60°C.

In the present invention, mRNA encoded by embryo-specific genes of the invention can be identified in Northern blots under stringent conditions using cDNAs of the invention or fragments of at least about 100 nucleotides. For the purposes of this disclosure, stringent conditions for such RNA-DNA hybridizations are those which include at least one wash in 0.2X SSC at 63°C for 20 minutes, or equivalent conditions. Genomic DNA or cDNA comprising genes of the invention can be identified using the same cDNAs (or fragments of at least about 100 nucleotides) under stringent conditions, which for purposes of this disclosure, include at least one wash (usually 2) in 0.2X SSC at a temperature of at least about 50°C, usually about 55°C, for 20 minutes, or equivalent conditions.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A shows a schematic representation of the three domains of the LEC1 polypeptide. Figure 2B shows a comparison of the predicted amino acid sequence of the B domain encoded by LEC1 with HAP3 homologs from maize, chicken, lamprey, *Xenopus* laevis, human, mouse, rat, *Emmericella nidulans*, *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae*, and *Kluyveromyces lactis*. The DNA-binding region and the subunit interaction region are indicated. Numbers indicate amino acid positions of the B domains.

DESCRIPTION OF THE PREFERRED EMBODIMENT

The present invention provides new embryo-specific genes useful in genetically engineering plants. Polynucleotide sequences from the genes of the invention can be used, for instance, to direct expression of desired heterologous genes in embryos (in the case of promoter sequences) or to modulate development of embryos or embryonic characteristics on other organs (e.g., by enhancing expression of the gene in a transgenic plant). In particular, the invention provides a new gene from *Arabidopsis* referred to here as LEC1. LEC1 encodes polypeptides which subunits of a protein which acts as a transcription factor. Thus, modulation of the expression of this gene can be used to manipulate a number of useful traits, such as increasing or decreasing storage protein content in cotyledons or leaves.

Generally, the nomenclature and the laboratory procedures in recombinant DNA technology described below are those well known and commonly employed in the art. Standard techniques are used for cloning, DNA and RNA isolation, amplification and purification. Generally enzymatic reactions involving DNA ligase, DNA polymerase, restriction endonucleases and the like are performed according to the manufacturer's specifications. These techniques and various other techniques are generally performed according to Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual*, 2nd. ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, (1989).

Isolation of nucleic acids of the invention

The isolation of sequences from the genes of the invention may be accomplished by a number of techniques. For instance, oligonucleotide probes based on the sequences disclosed here can be used to identify the desired gene in a cDNA or genomic DNA library from a desired plant species. To construct genomic libraries, large segments of genomic DNA are generated by random fragmentation, e.g. using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be packaged into the appropriate vector. To prepare a library of embryo-specific cDNAs, mRNA is isolated from embryos and a cDNA library that contains the gene transcripts is prepared from the mRNA.

The cDNA or genomic library can then be screened using a probe based upon the sequence of a cloned embryo-specific gene such as the polynucleotides disclosed here. Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species.

Alternatively, the nucleic acids of interest can be amplified from nucleic acid samples using amplification techniques. For instance, polymerase chain reaction (PCR) technology to amplify the sequences of the genes directly from mRNA, from cDNA, from genomic libraries or cDNA libraries. PCR and other in vitro amplification methods may also be useful, for example, to clone nucleic acid sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of the desired mRNA in samples, for nucleic acid sequencing, or for other purposes.

Appropriate primers and probes for identifying embryo-specific genes from plant tissues are generated from comparisons of the sequences provided herein. For a general overview of PCR see PCR Protocols: A Guide to Methods and Applications. (Innis, M, Gelfand, D., Sninsky, J. and White, T., eds.), Academic Press, San Diego (1990).

Appropriate primers for this purpose include, for instance: UP primer - 5' GGA ATT CAG CAA CAA CCC AAC CCC A 3' and LP primer - 5' LP primer - 5' GCT CTA GAC ATA CAA CAC TTT TCC TTA 3'. Alternatively, the following primer pairs can be used: 5' ATG ACC AGC TCA GTC ATA GTA GC 3' and 5' GCC ACA CAT GGT GGT TGC TGC TG 3' or 5' GAG ATA GAG ACC GAT CGT GGT TC 3' and 5' TCA CTT ATA CTG ACC ATA ATG GTC 3'. A third set of primers include: 5'-AGG ATC CAT GGA ACG TGG AGG CTT CCA T-3' and 5'-ATC TAG ATC AGT ACT TAT GTT GTT GAG TCG-3'. The amplifications conditions are typically as follows. Reaction components: 10 mM Tris-HCl, pH 8.3, 50 mM potassium chloride, 1.5 mM magnesium chloride, 0.001% gelatin, 200 microM (uM) dATP, 200 microM dCTP, 200 microM dGTP, 200 microM dTTP, 0.4 microM

Polynucleotides may also be synthesized by well-known techniques as described in the technical literature. See, e.g., Carruthers *et al.*, Cold Spring Harbor Symp.

Double stranded DNA fragments may then be obtained either by synthesizing the complementary strand and annealing the strands together under appropriate conditions, or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

The genus of LEC1 nucleic acid sequences of the invention includes genes and gene products identified and characterized by analysis using the sequences nucleic acid sequences, including SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:19 and SEQ ID NO:21, and protein sequences, including SEQ ID NO:2, SEQ ID NO:20 and SEQ ID NO:22. LEC1 sequences of the invention include nucleic acid sequences having substantial identity to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:19 and SEQ ID NO:21. LEC1 sequences of the invention include polypeptide sequences having substantial identify to SEQ ID NO:2, SEQ ID NO:20 and SEQ ID NO:22.

Gene fusions can be generated by basic recombinant DNA techniques as described below. Selection of gene fusions will depend on the desired phenotype caused by the gene fusion. For instance, if phenotypes associated with the A domain of one LEC1 protein are desired with phenotypes associated with the B domain of a second LEC1 protein, the a fusion of the first LEC1 protein's A domain to the second LEC1's B domain would be created. The fusion can subsequently be tested *in vitro* or *in vivo* for the desired phenotypes.

The isolated sequences prepared as described herein, can be used to prepare expression cassettes useful in a number of techniques. For example, expression cassettes of the invention can be used to suppress endogenous LEC1 gene expression. Inhibiting expression can be useful, for instance, in weed control (by transferring an inhibitory sequence

to a weedy species and allowing it to be transmitted through sexual crosses) or to produce fruit with small and non-viable seed.

A number of methods can be used to inhibit gene expression in plants. For instance, antisense technology can be conveniently used. To accomplish this, a nucleic acid segment from the desired gene is cloned and operably linked to a promoter such that the antisense strand of RNA will be transcribed. The expression cassette is then transformed into plants and the antisense strand of RNA is produced. In plant cells, it has been suggested that antisense RNA inhibits gene expression by preventing the accumulation of mRNA which encodes the enzyme of interest, see, e.g., Sheehy *et al.*, *Proc. Nat. Acad. Sci. USA*, 85:8805-8809 (1988), and Hiatt *et al.*, U.S. Patent No. 4,801,340.

The antisense nucleic acid sequence transformed into plants will be substantially identical to at least a portion of the endogenous embryo-specific gene or genes to be repressed. The sequence, however, does not have to be perfectly identical to inhibit expression. The vectors of the present invention can be designed such that the inhibitory effect applies to other proteins within a family of genes exhibiting homology or substantial homology to the target gene.

For antisense suppression, the introduced sequence also need not be full length relative to either the primary transcription product or fully processed mRNA. Generally, higher homology can be used to compensate for the use of a shorter sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and homology of non-coding segments may be equally effective. Normally, a sequence of between about 30 or 40 nucleotides and about full length nucleotides should be used, though a sequence of at least about 100 nucleotides is preferred, a sequence of at least about 200 nucleotides is more preferred, and a sequence of at least about 500 nucleotides is especially preferred.

Catalytic RNA molecules or ribozymes can also be used to inhibit expression of embryo-specific genes. It is possible to design ribozymes that specifically pair with virtually any target RNA and cleave the phosphodiester backbone at a specific location, thereby functionally inactivating the target RNA. In carrying out this cleavage, the ribozyme is not itself altered, and is thus capable of recycling and cleaving other molecules, making it a true enzyme. The inclusion of ribozyme sequences within antisense RNAs confers RNA-cleaving activity upon them, thereby increasing the activity of the constructs.

A number of classes of ribozymes have been identified. One class of ribozymes is derived from a number of small circular RNAs that are capable of self-cleavage and replication in plants. The RNAs replicate either alone (viroid RNAs) or with a helper

007020-2503T560

virus (satellite RNAs). Examples include RNAs from avocado sunblotch viroid and the satellite RNAs from tobacco ringspot virus, lucerne transient streak virus, velvet tobacco mottle virus, solanum nodiflorum mottle virus and subterranean clover mottle virus. The design and use of target RNA-specific ribozymes is described in Haseloff et al. *Nature*,
 5 334:585-591 (1988).

Another method of suppression is sense suppression. Introduction of expression cassettes in which a nucleic acid is configured in the sense orientation with respect to the promoter has been shown to be an effective means by which to block the transcription of target genes. For an example of the use of this method to modulate expression of
 10 endogenous genes see, Napoli et al., *The Plant Cell* 2:279-289 (1990), and U.S. Patents Nos. 5,034,323, 5,231,020, and 5,283,184.

Generally, where inhibition of expression is desired, some transcription of the introduced sequence occurs. The effect may occur where the introduced sequence contains no coding sequence per se, but only intron or untranslated sequences homologous to
 15 sequences present in the primary transcript of the endogenous sequence. The introduced sequence generally will be substantially identical to the endogenous sequence intended to be repressed. This minimal identity will typically be greater than about 65%, but a higher identity might exert a more effective repression of expression of the endogenous sequences. Substantially greater identity of more than about 80% is preferred, though about 95% to
 20 absolute identity would be most preferred. As with antisense regulation, the effect should apply to any other proteins within a similar family of genes exhibiting homology or substantial homology.

For sense suppression, the introduced sequence in the expression cassette, needing less than absolute identity, also need not be full length, relative to either the primary
 25 transcription product or fully processed mRNA. This may be preferred to avoid concurrent production of some plants which are overexpressers. A higher identity in a shorter than full length sequence compensates for a longer, less identical sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and identity of non-coding segments will be equally effective. Normally, a sequence of the size ranges noted
 30 above for antisense regulation is used.

One of skill in the art will recognize that using technology based on specific nucleotide sequences (*e.g.*, antisense or sense suppression technology), families of homologous genes can be suppressed with a single sense or antisense transcript. For instance, if a sense or antisense transcript is designed to have a sequence that is conserved

among a family of genes (*e.g.*, the B domain of LEC1), then multiple members of a gene family can be suppressed. Conversely, if the goal is to only suppress one member of a homologous gene family, then the sense or antisense transcript should be targeted to sequences with the most variance between family members. For instance, an antisense transcript identical to the A and C domains of LEC1 can be used to suppress LEC1 without suppressing related genes such as described in SEQ ID NO:19 or SEQ ID NO:21.

Another means of inhibiting LEC1 function in a plant is by creation of dominant negative mutations. In this approach, non-functional, mutant LEC1 polypeptides, which retain the ability to interact with wild-type subunits are introduced into a plant.

Identification of residues that can be changed to create a dominant negative can be determined by published work examining interaction of different subunits of CBF homologs from different species (see, *e.g.*, Sinha *et al.*, (1995). *Proc. Natl. Acad. Sci. USA* 92:1624-1628.)

Use of nucleic acids of the invention to enhance gene expression

Isolated sequences prepared as described herein can also be used to prepare expression cassettes which enhance or increase endogenous LEC1 gene expression. Where overexpression of a gene is desired, the desired gene from a different species may be used to decrease potential sense suppression effects. Enhanced expression of LEC1 polynucleotides is useful, for example, to increase storage protein content in plant tissues. Such techniques may be particularly useful for improving the nutritional value of plant tissues.

Any of a number of means well known in the art can be used to increase LEC1 activity in plants. Enhanced expression is useful, for example, to induce embryonic characteristics in plants or plant organs. Any organ can be targeted, such as shoot vegetative organs/structures (*e.g.* leaves, stems and tubers), roots, flowers and floral organs/structures (*e.g.* bracts, sepals, petals, stamens, carpels, anthers and ovules), seed (including embryo, endosperm, and seed coat) and fruit. Alternatively, one or several LEC1 genes can be expressed constitutively (*e.g.*, using the CaMV 35S promoter).

One of skill will recognize that the polypeptides encoded by the genes of the invention, like other proteins, have different domains which perform different functions. Thus, the gene sequences need not be full length, so long as the desired functional domain of the protein is expressed. As explained above, LEC1 polypeptides are related to CCAAT box-binding factor (CBF) proteins. CBFs are highly conserved family of transcription factors that regulate gene activity in eukaryotic organisms (see, *e.g.*, Mantvani (1992) *Nucl. Acids*

Res. 20:1087-1091; Li (1992) *Nucleic Acids Res.* 20:1087-1091). LEC1 was found to have high similarity to a portion of the HAP3 subunit of CBF. Thus, without being bound to any particular theory or mechanism, LEC1 is likely to act as a transcriptional modulator. HAP3 is divided into three domains, an amino terminal A domain, a central B domain, and a carboxyl terminal C domain, as shown diagrammatically in Figure 2A. Specifically, LEC1, has between about 75% and 85% sequence similarity, which is equivalent to 55% to 63% sequence identity, with the B domains of the other HAP3 homologs shown in Figure 2B; see also, Example 1, below. Figure 2B shows the amino acid sequence homology between LEC1 and other CBF homologs.

The LEC1 polypeptide also has an amino terminal A domain, a central B domain, and a carboxyl terminal C domain. The three domains of the LEC1 polypeptide are defined as follows: in SEQ ID NO:2, the A domain is located between about amino acid position 1 to about position 27; the B domain is located between about amino acid position 28 to about position 117; and, the C domain is located between about position 118 to about position 208. The B domain of LEC1, L1L and *Phaseolus* L1L are all closely related, whereas the A and C domains display almost no homology to each other.

The nucleotide sequence for LEC1 corresponding to each domain is displayed in SEQ ID NO 1, *e.g.*, the A domain is located between about nucleotide position 1 to about nucleotide position 82; the B domain is located between about nucleotide position 83 to about nucleotide position 351; the C domain is located between about nucleotide position 352 to about nucleotide position 624.

One of skill in the art will recognize that the domain boundaries are approximate. The boundaries for the domains of the LEC1 polypeptides and nucleotides can vary from 1 to 20 amino acids residues (1-60 nucleotides) from the boundaries listed above.

The DNA binding activity, and, therefore, transcription activation function, of LEC1 polypeptides is thought to be modulated by a short region of seven residues, MPIANVI (found, *e.g.*, at residues 34-40 of SEQ ID NO: 2). Thus, the polypeptides of the invention will often retain these sequences.

Modification of endogenous LEC1 genes

Methods for introducing genetic mutations into plant genes and selecting plants with desired traits are well known. For instance, seeds or other plant material can be treated with a mutagenic chemical substance, according to standard techniques. Such chemical substances include, but are not limited to, the following: diethyl sulfate, ethylene

imine, ethyl methanesulfonate and N-nitroso-N-ethylurea. Alternatively, ionizing radiation from sources such as, X-rays or gamma rays can be used.

Modified protein chains can also be readily designed utilizing various recombinant DNA techniques well known to those skilled in the art and described for instance, in Sambrook et al., *supra*. Hydroxylamine can also be used to introduce single base mutations into the coding region of the gene (Sikorski, *et al.*, (1991). *Meth. Enzymol.* 194: 302-318). For example, the chains can vary from the naturally occurring sequence at the primary structure level by amino acid substitutions, additions, deletions, and the like. These modifications can be used in a number of combinations to produce the final modified protein chain.

Alternatively, homologous recombination can be used to induce targeted gene modifications by specifically targeting the *LEC1* gene *in vivo* (*see, generally*, Grewal and Klar, *Genetics* 146: 1221-1238 (1997) and Xu *et al.*, *Genes Dev.* 10: 2411-2422 (1996)). Homologous recombination has been demonstrated in plants (Puchta *et al.*, *Experientia* 50: 277-284 (1994), Swoboda *et al.*, *EMBO J.* 13: 484-489 (1994); Offringa *et al.*, *Proc. Natl. Acad. Sci. USA* 90: 7346-7350 (1993); and Kempin *et al.* *Nature* 389:802-803 (1997)).

In applying homologous recombination technology to the genes of the invention, mutations in selected portions of an *LEC1* gene sequences (including 5' upstream, 3' downstream, and intragenic regions) such as those disclosed here are made *in vitro* and then introduced into the desired plant using standard techniques. Since the efficiency of homologous recombination is known to be dependent on the vectors used, use of dicistronic gene targeting vectors as described by Mountford *et al.*, *Proc. Natl. Acad. Sci. USA* 91: 4303-4307 (1994); and Vaultont *et al.*, *Transgenic Res.* 4: 247-255 (1995) are conveniently used to increase the efficiency of selecting for altered *LEC1* gene expression in transgenic plants.

The mutated gene will interact with the target wild-type gene in such a way that homologous recombination and targeted replacement of the wild-type gene will occur in transgenic plant cells, resulting in suppression of *LEC1* activity.

Alternatively, oligonucleotides composed of a contiguous stretch of RNA and DNA residues in a duplex conformation with double hairpin caps on the ends can be used.

The RNA/DNA sequence is designed to align with the sequence of the target *LEC1* gene and to contain the desired nucleotide change. Introduction of the chimeric oligonucleotide on an extrachromosomal T-DNA plasmid results in efficient and specific *LEC1* gene conversion directed by chimeric molecules in a small number of transformed plant cells. This method is

described in Cole-Strauss *et al.*, *Science* 273:1386-1389 (1996) and Yoon *et al.* *Proc. Natl. Acad. Sci. USA* 93: 2071-2076 (1996).

Desired modified LEC1 polypeptides can be identified using assays to screen for the presence or absence of wild type LEC1 activity. Such assays can be based on the ability of the LEC1 protein to functionally complement the hap3 mutation in yeast. As noted above, it has been shown that homologs from different species functionally interact with yeast subunits of the CBF. (Sinha, *et al.*, (1995). *Proc. Natl. Acad. Sci. USA* 92:1624-1628); see, also, Becker, *et al.*, (1991). *Proc. Natl. Acad. Sci. USA* 88: 1968-1972). The reporter for this screen can be any of a number of standard reporter genes such as the lacZ gene encoding beta-galactosidase that is fused with the regulatory DNA sequences and promoter of the yeast CYC1 gene. This promoter is regulated by the yeast CBF.

A plasmid containing the LEC1 cDNA clone is mutagenized in vitro according to techniques well known in the art. The cDNA inserts are excised from the plasmid and inserted into the cloning site of a yeast expression vector such as pYES2 (Invitrogen). The plasmid is introduced into hap3- yeast containing a lacZ reporter that is regulated by the yeast CBF such as pLG265UP1-lacZ (Guarente, *et al.*, (1984) *Cell* 36: 317-321). Transformants are then selected and a filter assay is used to test colonies for beta-galactosidase activity. After confirming the results of activity assays, immunochemical tests using a LEC1 antibody are performed on yeast lines that lack beta-galactosidase activity to identify those that produce stable LEC1 protein but lack activity. The mutant LEC1 genes are then cloned from the yeast and their nucleotide sequence determined to identify the nature of the lesions.

In other embodiments, the promoters derived from the LEC1 genes of the invention can be used to drive expression of heterologous genes in an embryo-specific or seed-specific manner, such that desired gene products are present in the embryo, seed, or fruit. Suitable structural genes that could be used for this purpose include genes encoding proteins useful in increasing the nutritional value of seed or fruit. Examples include genes encoding enzymes involved in the biosynthesis of antioxidants such as vitamin A, vitamin C, vitamin E and melatonin. Other suitable genes encoding proteins involved in modification of fatty acids, or in the biosynthesis of lipids, proteins, and carbohydrates. Still other genes can be those encoding proteins involved in auxin and auxin analog biosynthesis for increasing fruit size, genes encoding pharmaceutically useful compounds, and genes encoding plant resistance products to combat fungal or other infections of the seed.

Typically, desired promoters are identified by analyzing the 5' sequences of a genomic clone corresponding to the embryo-specific genes described here. Sequences characteristic of promoter sequences can be used to identify the promoter. Sequences controlling eukaryotic gene expression have been extensively studied. For instance, promoter sequence elements include the TATA box consensus sequence (TATAAT), which is usually 20 to 30 base pairs upstream of the transcription start site. In most instances the TATA box is required for accurate transcription initiation. In plants, further upstream from the TATA box, at positions -80 to -100, there is typically a promoter element with a series of adenines surrounding the trinucleotide G (or T) N G. J. Messing et al., in *Genetic Engineering in Plants*, pp. 221-227 (Kosage, Meredith and Hollaender, eds. (1983)).

A number of methods are known to those of skill in the art for identifying and characterizing promoter regions in plant genomic DNA (see, e.g., Jordano, *et al.*, *Plant Cell*, 1: 855-866 (1989); Bustos, *et al.*, *Plant Cell*, 1:839-854 (1989); Green, *et al.*, *EMBO J.* 7, 4035-4044 (1988); Meier, *et al.*, *Plant Cell*, 3, 309-316 (1991); and Zhang, *et al.*, *Plant Physiology* 110: 1069-1079 (1996)).

Preparation of recombinant vectors

To use isolated sequences in the above techniques, recombinant DNA vectors suitable for transformation of plant cells are prepared. Techniques for transforming a wide variety of higher plant species are well known and described in the technical and scientific literature. See, for example, Weising *et al.* *Ann. Rev. Genet.* 22:421-477 (1988). A DNA sequence coding for the desired polypeptide, for example a cDNA sequence encoding a full length protein, will preferably be combined with transcriptional and translational initiation regulatory sequences which will direct the transcription of the sequence from the gene in the intended tissues of the transformed plant.

For example, for overexpression, a plant promoter fragment may be employed which will direct expression of the gene in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens*, and other transcription initiation regions from various plant genes known to those of skill.

Alternatively, the plant promoter may direct expression of the polynucleotide of the invention in a specific tissue (tissue-specific promoters) or may be otherwise under

more precise environmental control (inducible promoters). Examples of tissue-specific promoters under developmental control include promoters that initiate transcription only in certain tissues, such as fruit, seeds, or flowers. As noted above, the promoters from the LEC1 genes described here are particularly useful for directing gene expression so that a desired gene product is located in embryos or seeds. Other suitable promoters include those from genes encoding storage proteins or the lipid body membrane protein, oleosin. Examples of environmental conditions that may affect transcription by inducible promoters include anaerobic conditions, elevated temperature, or the presence of light.

If proper polypeptide expression is desired, a polyadenylation region at the 3'-end of the coding region should be included. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA.

The vector comprising the sequences (e.g., promoters or coding regions) from genes of the invention will typically comprise a marker gene which confers a selectable phenotype on plant cells. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosulfuron or Basta.

LEC1 nucleic acid sequences of the invention are expressed recombinantly in plant cells to enhance and increase levels of endogenous LEC1 polypeptides. Alternatively, antisense or other LEC1 constructs (described above) are used to suppress LEC1 levels of expression. A variety of different expression constructs, such as expression cassettes and vectors suitable for transformation of plant cells can be prepared. Techniques for transforming a wide variety of higher plant species are well known and described in the technical and scientific literature. See, e.g., Weising *et al. Ann. Rev. Genet.* 22:421-477 (1988). A DNA sequence coding for a LEC1 polypeptide, e.g., a cDNA sequence encoding a full length protein, can be combined with cis-acting (promoter) and trans-acting (enhancer) transcriptional regulatory sequences to direct the timing, tissue type and levels of transcription in the intended tissues of the transformed plant. Translational control elements can also be used.

The invention provides a LEC1 nucleic acid operably linked to a promoter which, in a preferred embodiment, is capable of driving the transcription of the LEC1 coding sequence in plants. The promoter can be, e.g., derived from plant or viral sources. The promoter can be, e.g., constitutively active, inducible, or tissue specific. In construction of recombinant expression cassettes, vectors, transgenics, of the invention, a different promoters

can be chosen and employed to differentially direct gene expression, e.g., in some or all tissues of a plant or animal.

Typically, desired promoters are identified by analyzing the 5' sequences of a genomic clone corresponding to the embryo-specific genes described here. Sequences characteristic of promoter sequences can be used to identify the promoter. Sequences controlling eukaryotic gene expression have been extensively studied. For instance, promoter sequence elements include the TATA box consensus sequence (TATAAT), which is usually 20 to 30 base pairs upstream of the transcription start site. In most instances the TATA box is required for accurate transcription initiation. In plants, further upstream from the TATA box, at positions -80 to -100, there is typically a promoter element with a series of adenines surrounding the trinucleotide G (or T) N G. J. Messing et al., in *Genetic Engineering in Plants*, pp. 221-227 (Kosage, Meredith and Hollaender, eds. (1983)). A number of methods are known to those of skill in the art for identifying and characterizing promoter regions in plant genomic DNA (see, e.g., Jordano, *et al.*, *Plant Cell*, 1: 855-866 (1989); Bustos, *et al.*, *Plant Cell*, 1:839-854 (1989); Green, *et al.*, *EMBO J.* 7, 4035-4044 (1988); Meier, *et al.*, *Plant Cell*, 3, 309-316 (1991); and Zhang (1996) *Plant Physiology* 110:1069-1079).

Constitutive Promoters

A promoter fragment can be employed which will direct expression of LEC1 nucleic acid in all transformed cells or tissues, e.g. as those of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Promoters that drive expression continuously under physiological conditions are referred to as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include those from viruses which infect plants, such as the cauliflower mosaic virus (CaMV) 35S transcription initiation region (see, e.g., Dagless (1997) *Arch. Virol.* 142:183-191); the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens* (see, e.g., Mengiste (1997) *supra*; O'Grady (1995) *Plant Mol. Biol.* 29:99-108); the promoter of the tobacco mosaic virus; the promoter of Figwort mosaic virus (see, e.g., Maiti (1997) *Transgenic Res.* 6:143-156); actin promoters, such as the *Arabidopsis* actin gene promoter (see, e.g., Huang (1997) *Plant Mol. Biol.* 1997 33:125-139); alcohol dehydrogenase (Adh) gene promoters (see, e.g., Millar (1996) *Plant Mol. Biol.* 31:897-904); *ACT11* from *Arabidopsis* (Huang *et al.* *Plant Mol. Biol.* 33:125-139 (1996)), *Cat3* from *Arabidopsis* (GenBank No. U43147, Zhong *et al.*, *Mol. Gen. Genet.* 251:196-203 (1996)), the gene encoding stearyl-acyl carrier protein desaturase from *Brassica napus*

(Genbank No. X74782, Solocombe *et al.* *Plant Physiol.* 104:1167-1176 (1994)), *Gpc1* from maize (GenBank No. X15596, Martinez *et al.* *J. Mol. Biol.* 208:551-565 (1989)), *Gpc2* from maize (GenBank No. U45855, Manjunath *et al.*, *Plant Mol. Biol.* 33:97-112 (1997)), other transcription initiation regions from various plant genes known to those of skill. See also

5 Holtorf (1995) "Comparison of different constitutive and inducible promoters for the overexpression of transgenes in *Arabidopsis thaliana*," *Plant Mol. Biol.* 29:637-646.

Inducible Promoters

Alternatively, a plant promoter may direct expression of the LEC1 nucleic acid of the invention under the influence of changing environmental conditions or

10 developmental conditions. Examples of environmental conditions that may effect transcription by inducible promoters include anaerobic conditions, elevated temperature, drought, or the presence of light. Such promoters are referred to herein as "inducible" promoters. For example, the invention incorporates the drought-inducible promoter of maize (Busk (1997) *supra*); the cold, drought, and high salt inducible promoter from potato (Kirch

15 (1997) *Plant Mol. Biol.* 33:897-909).

Alternatively, plant promoters which are inducible upon exposure to plant hormones, such as auxins, are used to express the nucleic acids of the invention. For example, the invention can use the auxin-response elements E1 promoter fragment (AuxREs) in the soybean (*Glycine max* L.) (Liu (1997) *Plant Physiol.* 115:397-407); the auxin-

20 responsive *Arabidopsis* GST6 promoter (also responsive to salicylic acid and hydrogen peroxide) (Chen (1996) *Plant J.* 10: 955-966); the auxin-inducible *parC* promoter from tobacco (Sakai (1996) 37:906-913); a plant biotin response element (Streit (1997) *Mol. Plant Microbe Interact.* 10:933-937); and, the promoter responsive to the stress hormone abscisic acid (Sheen (1996) *Science* 274:1900-1902).

Plant promoters which are inducible upon exposure to chemicals reagents which can be applied to the plant, such as herbicides or antibiotics, are also used to express the nucleic acids of the invention. For example, the maize In2-2 promoter, activated by benzenesulfonamide herbicide safeners, can be used (De Veylder (1997) *Plant Cell Physiol.* 38:568-577); application of different herbicide safeners induces distinct gene expression

30 patterns, including expression in the root, hydathodes, and the shoot apical meristem. LEC1 coding sequence can also be under the control of, e.g., a tetracycline-inducible promoter, e.g., as described with transgenic tobacco plants containing the *Avena sativa* L. (oat) arginine decarboxylase gene (Masgrau (1997) *Plant J.* 11:465-473); or, a salicylic acid-responsive element (Stange (1997) *Plant J.* 11:1315-1324).

Tissue-Specific Promoters

Alternatively, the plant promoter may direct expression of the polynucleotide of the invention in a specific tissue (tissue-specific promoters). Tissue specific promoters are transcriptional control elements that are only active in particular cells or tissues at specific times during plant development, such as in vegetative tissues or reproductive tissues.

Promoters from the LEC1 genes of the invention are particularly useful for tissue-specific direction of gene expression so that a desired gene product is generated only or preferentially in embryos or seeds, as described below.

Examples of tissue-specific promoters under developmental control include promoters that initiate transcription only (or primarily only) in certain tissues, such as vegetative tissues, e.g., roots or leaves, or reproductive tissues, such as fruit, ovules, seeds, pollen, pistils, flowers, or any embryonic tissue. Reproductive tissue-specific promoters may be, e.g., ovule-specific, embryo-specific, endosperm-specific, integument-specific, seed and seed coat-specific, pollen-specific, petal-specific, sepal-specific, or some combination thereof.

Suitable seed-specific promoters are derived from the following genes: MAC1 from maize, Sheridan (1996) *Genetics* 142:1009-1020; Cat3 from maize, GenBank No. L05934, Abler (1993) *Plant Mol. Biol.* 22:10131-1038; viviparous-1 from Arabidopsis, Genbank No. U93215; atmyc1 from Arabidopsis, Urao (1996) *Plant Mol. Biol.* 32:571-57; Conceicao (1994) *Plant* 5:493-505; napA from Brassica napus, GenBank No. J02798, Josefsson (1987) *JBL* 26:12196-1301; the napin gene family from Brassica napus, Sjobahl (1995) *Planta* 197:264-271.

The ovule-specific BEL1 gene described in Reiser (1995) *Cell* 83:735-742, GenBank No. U39944, can also be used. See also Ray (1994) *Proc. Natl. Acad. Sci. USA* 91:5761-5765. The egg and central cell specific FIE1 promoter is also a useful reproductive tissue-specific promoter.

Sepal and petal specific promoters are also used to express LEC1 nucleic acids in a reproductive tissue-specific manner. For example, the Arabidopsis floral homeotic gene APETALA1 (AP1) encodes a putative transcription factor that is expressed in young flower primordia, and later becomes localized to sepals and petals (see, e.g., Gustafson- Brown (1994) *Cell* 76:131-143; Mandel (1992) *Nature* 360:273-277). A related promoter, for AP2, a floral homeotic gene that is necessary for the normal development of sepals and petals in floral whorls, is also useful (see, e.g., Drews (1991) *Cell* 65:991-1002; Bowman (1991) *Plant Cell* 3:749-758). Another useful promoter is that controlling the expression of the unusual

floral organs (ufo) gene of Arabidopsis, whose expression is restricted to the junction between sepal and petal primordia (Bossinger (1996) *Development* 122:1093-1102).

A maize pollen-specific promoter has been identified in maize (Guerrero (1990) *Mol. Gen. Genet.* 224:161-168). Other genes specifically expressed in pollen are described, e.g., by Wakeley (1998) *Plant Mol. Biol.* 37:187-192; Ficker (1998) *Mol. Gen. Genet.* 257:132-142; Kulikauskas (1997) *Plant Mol. Biol.* 34:809-814; Treacy (1997) *Plant Mol. Biol.* 34:603-611.

Other suitable promoters include those from genes encoding embryonic storage proteins. For example, the gene encoding the 2S storage protein from Brassica napus, Dasgupta (1993) *Gene* 133:301-302; the 2s seed storage protein gene family from Arabidopsis; the gene encoding oleosin 20kD from Brassica napus, GenBank No. M63985; the genes encoding oleosin A, Genbank No. U09118, and, oleosin B, Genbank No. U09119, from soybean; the gene encoding oleosin from Arabidopsis, Genbank No. Z17657; the gene encoding oleosin 18kD from maize, GenBank No. J05212, Lee (1994) *Plant Mol. Biol.* 26:1981-1987; and, the gene encoding low molecular weight sulphur rich protein from soybean, Choi (1995) *Mol Gen. Genet.* 246:266-268, can be used. The tissue specific E8 promoter from tomato is particularly useful for directing gene expression so that a desired gene product is located in fruits.

A tomato promoter active during fruit ripening, senescence and abscission of leaves and, to a lesser extent, of flowers can be used (Blume (1997) *Plant J.* 12:731-746). Other exemplary promoters include the pistil specific promoter in the potato (*Solanum tuberosum* L.) SK2 gene, encoding a pistil-specific basic endochitinase (Ficker (1997) *Plant Mol. Biol.* 35:425-431); the Blec4 gene from pea (*Pisum sativum* cv. Alaska), active in epidermal tissue of vegetative and floral shoot apices of transgenic alfalfa. This makes it a useful tool to target the expression of foreign genes to the epidermal layer of actively growing shoots.

A variety of promoters specifically active in vegetative tissues, such as leaves, stems, roots and tubers, can also be used to express the LEC1 nucleic acids of the invention. For example, promoters controlling patatin, the major storage protein of the potato tuber, can be used, see, e.g., Kim (1994) *Plant Mol. Biol.* 26:603-615; Martin (1997) *Plant J.* 11:53-62. The ORF13 promoter from *Agrobacterium rhizogenes* which exhibits high activity in roots can also be used (Hansen (1997) *Mol. Gen. Genet.* 254:337-343). Other useful vegetative tissue-specific promoters include: the tarin promoter of the gene encoding a globulin from a major taro (*Colocasia esculenta* L. Schott) corm protein family, tarin (Bezerra (1995) *Plant*

Mol. Biol. 28:137-144); the curculin promoter active during taro corm development (de Castro (1992) *Plant Cell* 4:1549-1559) and the promoter for the tobacco root-specific gene TobRB7, whose expression is localized to root meristem and immature central cylinder regions (Yamamoto (1991) *Plant Cell* 3:371-382).

5 Leaf-specific promoters, such as the ribulose biphosphate carboxylase (RBCS) promoters can be used. For example, the tomato RBCS1, RBCS2 and RBCS3A genes are expressed in leaves and light-grown seedlings, only RBCS1 and RBCS2 are expressed in developing tomato fruits (Meier (1997) *FEBS Lett.* 415:91-95). A ribulose bisphosphate carboxylase promoters expressed almost exclusively in mesophyll cells in leaf blades and leaf
10 sheaths at high levels, described by Matsuoka (1994) *Plant J.* 6:311-319, can be used. Another leaf-specific promoter is the light harvesting chlorophyll a/b binding protein gene promoter, see, e.g., Shiina (1997) *Plant Physiol.* 115:477-483; Casal (1998) *Plant Physiol.* 116:1533-1538. The Arabidopsis thaliana myb-related gene promoter (Atmyb5) described by Li (1996) *FEBS Lett.* 379:117-121, is leaf-specific. The Atmyb5 promoter is expressed in
15 developing leaf trichomes, stipules, and epidermal cells on the margins of young rosette and cauline leaves, and in immature seeds. Atmyb5 mRNA appears between fertilization and the 16 cell stage of embryo development and persists beyond the heart stage. A leaf promoter identified in maize by Busk (1997) *Plant J.* 11:1285-1295, can also be used.

Another class of useful vegetative tissue-specific promoters are meristematic
20 (root tip and shoot apex) promoters. For example, the "SHOOTMERISTEMLESS" and "SCARECROW" promoters, which are active in the developing shoot or root apical meristems, described by Di Laurenzio (1996) *Cell* 86:423-433; and, Long (1996) *Nature* 379:66-69; can be used. Another useful promoter is that which controls the expression of 3-hydroxy-3- methylglutaryl coenzyme A reductase HMG2 gene, whose expression is
25 restricted to meristematic and floral (secretory zone of the stigma, mature pollen grains, gynoecium vascular tissue, and fertilized ovules) tissues (see, e.g., Enjuto (1995) *Plant Cell.* 7:517-527). Also useful are kn1-related genes from maize and other species which show meristem-specific expression, see, e.g., Granger (1996) *Plant Mol. Biol.* 31:373-378; Kerstetter (1994) *Plant Cell* 6:1877-1887; Hake (1995) *Philos. Trans. R. Soc. Lond. B. Biol.*
30 *Sci.* 350:45-51. For example, the Arabidopsis thaliana KNAT1 promoter. In the shoot apex, KNAT1 transcript is localized primarily to the shoot apical meristem; the expression of KNAT1 in the shoot meristem decreases during the floral transition and is restricted to the cortex of the inflorescence stem (see, e.g., Lincoln (1994) *Plant Cell* 6:1859-1876).

5

10

15

20

30

Agrobacterium tumefaciens-mediated transformation techniques, including disarming and use of binary vectors, are well described in the scientific literature. See, for

example Horsch *et al. Science* 233:496-498 (1984), and Fraley *et al. Proc. Natl. Acad. Sci. USA* 80:4803 (1983).

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype and thus the desired phenotype such as seedlessness. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans *et al.*, *Protoplasts Isolation and Culture*, Handbook of Plant Cell Culture, pp. 124-176, MacMillan Publishing Company, New York, 1983; and Binding, *Regeneration of Plants, Plant Protoplasts*, pp. 21-73, CRC Press, Boca Raton, 1985. Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee *et al. Ann. Rev. of Plant Phys.* 38:467-486 (1987).

The nucleic acids of the invention can be used to confer desired traits on essentially any plant. Thus, the invention has use over a broad range of plants, including species from the genera Asparagus, Atropa, Avena, Brassica, Citrus, Citrullus, Capsicum, Cucumis, Cucurbita, Daucus, Fragaria, Glycine, Gossypium, Helianthus, Heterocallis, Hordeum, Hyoscyamus, Lactuca, Linum, Lolium, Lycopersicon, Malus, Manihot, Majorana, Medicago, Nicotiana, Oryza, Panieum, Pannasetum, Persea, Pisum, Pyrus, Prunus, Raphanus, Secale, Senecio, Sinapis, Solanum, Sorghum, Trigonella, Triticum, Vitis, Vigna, and, Zea. The LEC1 genes of the invention are particularly useful in the production of transgenic plants in the genus Brassica. Examples include broccoli, cauliflower, brussel sprouts, canola, and the like.

Use and Recombinant Expression of LEC1 in Combination with other Genes

The LEC1 nucleic acids of the invention can be expressed together with other structural or regulatory genes to achieve a desired effect. A cell or plant, such as a transformed cell or a transgenic plant, can be transformed, engineered or bred to co-express both LEC1 nucleotide and/or LEC1 polypeptide, and another gene or gene product. Alternatively, two or more LEC1 nucleic acids can be co-expressed together in the same plant or cell.

The LEC1 nucleic acids of the invention, when expressed in plant reproductive or vegetative tissue, can induce ectopic embryo morphogenesis. Thus, in one embodiment, a LEC1 nucleic acid of the invention is expressed in a sense conformation in a

transgenic plant to induce the expression of ectopic embryo-like structures, as discussed above. In another embodiment, LEC1 is co-expressed with a gene or nucleic acid that increases reproductive tissue mass, e.g., increases fruit size, seed mass, seed protein or seed oils. For example, co-expression of antisense nucleic acid to ADC genes, such as AP2 and RAP2 genes of Arabidopsis, will dramatically increase seed mass, seed protein and seed oils; see, e.g., Jofuku, *et al.*, WO 98/07842; Okamuro (1997) *Proc. Natl. Acad. Sci. USA* 94:7076-7081; Okamuro (1997) *Plant Cell* 9:37-47; Jofuku (1994) *Plant Cell* 6:1211-1225. Thus, co-expression of a LEC1 of the invention, to induce ectopic expression of embryonic cells and tissues, together with another plant nucleic acid and/or protein, such as the seed-mass enhancing antisense AP2 nucleic acid, generates a cell, tissue, or plant (e.g., a transgenic plant) with increased fruit and seed mass, greater yields of embryonic storage proteins, and the like.

In another embodiment, the LEC1 nucleic acids of the invention are expressed in plant reproductive or vegetative cells and tissues which lack the ability to produce functional ADC genes, such as AP2 and RAP2 genes. The LEC1 nucleic acid can be expressed in an ADC "knockout" transgenic plant. Alternatively, the LEC1 nucleic acid can be expressed in a cell, tissue or plant expressing a mutant ADC nucleic acid or gene product. Expression of LEC1 nucleic acid in any of these non-functioning ADC models will also produce a cell, tissue or plant with increased fruit and seed mass, greater yields of embryonic storage proteins, and the like.

One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

Example 1

This example describes the isolation and characterization of an exemplary LEC1 gene.

5 Experimental Procedures

Plant Material

A lec1-2 mutant was identified from a population of *Arabidopsis thaliana* ecotype Wassilewskija (Ws-O) lines mutagenized with T-DNA insertions as described before (West et al., 1994). The abi3-3, fus3-3 and lec1-1 mutants were generously provided by
10 Peter McCourt, University of Toronto and David Meinke, Oklahoma State University. Wild type plants and mutants were grown under constant light at 22°C.

Double mutants were constructed by intercrossing the mutant lines lec1-1, lec1-2, abi3-3, fus3-3, and lec2. The genotype of the double mutants was verified through backcrosses with each parental line. Double mutants were those who failed to complement
15 both parent lines. Homozygous single and double mutants were generated by germinating intact seeds or dissected mature embryos before desiccation on basal media.

Isolation and Sequence analysis of Genomic and cDNA Clones

Genomic libraries of Ws-O wild type plants, lec1-1 and lec1-2 mutants were made in GEM11 vector according to the instructions of the manufacturer (Promega). Two
20 silique-specific cDNA libraries (stages globular to heart and heart to young torpedo) were made in ZAPII vector (Stratagene).

The genomic library of lec1-2 was screened using right and left T-DNA specific probes according to standard techniques. About 12 clones that cosegregate with the mutation, were isolated and purified and the entire DNAs were further labeled and used as
25 probes to screen a southern blot containing wild type and lec1-1 genomic DNA. One clone hybridized with plant DNA and was further analyzed. A 7.1 kb XhoI fragment containing the left border and the plant sequence flanking the T-DNA was subcloned into pBluescript-KS plasmid (Stratagene) to form ML7 and sequenced using a left border specific primer (5' GCATAGATGCACTCGAAATCAGCC 3'). The T-DNA organization was
30 partially verified using southern analysis with T-DNA left and right borders and PBR322 probes. The results suggested that the other end of the T-DNA is also composed of left border. This was confirmed by generating a PCR fragment using a genomic plant DNA primer (LP primer 5' GCT CTA GAC ATA CAA CAC TTT TCC TTA 3') and a T-DNA left border specific primer (5' GCTTGGTAATAATTGTCATTAG 3') and sequencing.

The EcoRI insert of ML7 was used to screen a wild type genomic library. Two overlapping clones were purified and a 7.4 EcoRI genomic fragment from the wild type DNA region was subcloned into pBluescript-KS plasmid making WT74. This fragment was sequenced (SEQ ID NO: 4) and was used to screen lec1-1 genomic library and wild type
 5 silique-specific cDNA libraries. 8 clones from the lec1-1 genomic library were identified and analyzed by restriction mapping.

From these clones the exact site of the deletion in lec1-1 was mapped and sequenced by amplifying a Xbp PCR fragment using primers (H21 - 5' H21 - 5' CTA AAA ACA TCT ACG GTT CA 3'; H 17 - 5' TTT GTG GTT GAC CGT TTG GC 3') flanking the
 10 deletion region in lec1-1 genomic DNA. Clones were isolated from both cDNA libraries and partially sequenced. The sequence of the cDNA clones and the wild type genomic clone matched exactly, confirming that both derived from the same locus. All hybridizations were performed under stringent conditions with 32P random prime probes (Stratagene).

Sequencing was done using the automated dideoxy chain termination method
 15 (Applied Biosystems, Foster City, CA). Data base searches were performed at the National Center for Biotechnology Information by using the BLAST network service. Alignment of protein sequences was done using PILEUP program (Genetics Computer Group, Madison, WI)

DNA and RNA blot analysis

20 Genomic DNA was isolated from leaves by using the CTAB-containing buffer Dellaporta, et al., (1983). Plant Mol. Biol. Reporter 1: 19-21. Two micrograms of DNA was digested with different restriction endonucleases, electrophoretically separated in 1% agarose gel, and transferred to a nylon membrane (Hybond N; Amersham).

Total RNA was prepared from siliques, two days old seedlings, stems, leaves,
 25 buds and roots. Poly(A)+ RNA was purified from total RNA by oligo(dT) cellulose chromatography, and two micrograms of each Poly(A)+ RNA samples were separated in 1% denatured formaldehyde-agarose gel. Hybridizations were done under stringent conditions unless it specifies otherwise. Radioactive probes were prepared as described above.

Complementation of lec1 mutants

A 3.4 kb BstYI fragment of genomic DNA (SEQ ID NO: 3) containing sequences from 1.992 kb upstream of the ORF to a region 579 bp downstream from the poly A site was subcloned into the hygromycin resistant binary vector pBIB-Hyg. The LEC1 cDNA was placed under the control of the 35S promoter and the ocs polyadenylation signals by inserting a PCR fragment spanning the entire coding region into the plasmid pART7. The entire regulatory fragment was then removed by digestion with NotI and transferred into the hygromycin resistant binary vector BJ49. The binary vectors were introduced into the Agrobacterium strain GV3101, and constructions were checked by re-isolation of the plasmids and restriction enzyme mapping, or by PCR. Transformation to homozygous lec1-1 and lec1-2 mutants were done using the in planta transformation procedure (Bechtold, et al., (1993). Comptes Rendus de l'Academie des Sciences Serie III Sciences de la Vie, 316: 1194-1199. Dry seeds from lec1 mutants were selected for transformants by their ability to germinate after desiccation on plates containing 5g/ml hygromycin. The transformed plants were tested for the present of the transgene by PCR and by screening the siliques for the present of viable seeds.

In Situ Hybridization

Experiments were performed as described previously by Dietrich et al. (1989) Plant Cell 1: 73-80. Sections were hybridized with LEC1 antisense probe. As a negative control, the LEC1 antisense probe was hybridized to seed sections of lec1 mutants. In addition, a sense probe was prepared and reacted with the wild type seed sections.

Results

Genetic Interaction Between Leafy Cotyledon-Type Mutants and abi3

In order to understand the genetic pathways which regulate late embryogenesis we took advantage of three Arabidopsis mutants lec2, fus3-3 and abi3-3 that cause similar defects in late embryogenesis to those of lec1-1 or lec1-2. These mutants are desiccation intolerant, sometimes viviparous and have activated shoot apical meristems. The lec2 and fus3-3 mutants are sensitive to ABA and possess trichomes on their cotyledons and therefore can be categorized as leafy cotyledon-type mutants (Meinke et al., 1994). The abi3-3 mutants belong to a different class of late embryo defective mutations that is insensitive to ABA and does not have trichomes on the cotyledons.

The two classes of mutants were crossed to lec1-1 and lec1-2 mutants to construct plants homozygous to both mutations. The lec1 and lec2 mutations interact

synergistically, resulting in a double mutant which is arrested in a stage similar to the late heart stage, the double mutant embryo, however, is larger. The *lec1* or *lec2* and *fus3-3* double mutants did not display any epistasis and the resulting embryo had an intermediate phenotype. The *lec1/abi3-3* double mutants and *lec2/abi3-3* double mutants were ABA insensitive and had a *lec*-like phenotype. There was no difference between double mutants that consist of either *lec1-1* or *lec1-2*.

No epistasis was seen between the double mutants indicating that each of the above genes, the LEC-type and ABI3 genes, operate in different genetic pathways.

LEC1 Functions Early in Embryogenesis

The effects of *lec1* is not limited to late embryogenesis, it also has a role in early embryogenesis. The embryos of the *lec1/lec2* double mutants were arrested in the early stages of development, while the single mutants developed into mature embryos, suggesting that these genes act early during development.

Further examination of the early stages of the single and double mutations showed defects in the shape, size and cell division pattern of the mutants suspensors. The suspensor of wild type embryo consists of a single file of six to eight cells, whereas the suspensors of the mutants are often enlarged and undergo periclinal divisions. Leafy cotyledon mutants exhibit suspensor anomalies at the globular or transition stage whereas wild type and *abi3* mutant do not show any abnormalities.

The number of anomalous suspensors increases as the embryos continue to develop. At the torpedo stage, the wild type suspensor cells undergo programmed cell death, but in the mutants secondary embryos often develop from the abnormal suspensors and, when rescued, give rise to twins.

The Organization of the LEC1 Locus in Wild Type Plants and *lec1* Mutants

Two mutant alleles of the LEC1 gene have been reported, *lec1-1* and *lec1-2* (Meinke, 1992; West et al., 1994). Both mutants were derived from a population of plants mutagenized insertionally with T-DNA (Feldmann and Marks, 1987), although *lec1-1* is not tagged. The *lec1-2* mutant contains multiple T-DNA insertions. A specific subset of T-DNA fragments were found to be closely linked with the mutation. A genomic library of *lec1-2* was screened using right and left border T-DNA as probes. Genomic clones containing T-DNA fragments that cosegregate with the mutation were isolated and tested on Southern blots of both wild type and *lec1-1* plants. Only one clone hybridized with Arabidopsis DNA and also gave polymorphic restriction fragment in *lec1-1*.

The lec1-1 polymorphism resulted from a small deletion, approximately 2 kb in length. Using sequences from the plant fragment flanking the T-DNA, the genomic wild type DNA clones and the lec1-1 genomic clones were isolated. An EcoRI fragment of 7.4 kb of the genomic wild type DNA that corresponded to the polymorphic restriction fragment in lec1-1 was further analyzed and sequenced. The exact site of the deletion in lec1-1 was identified using a PCR fragment that was generated by primers, within the expected borders of the deleted fragment, and sequencing.

In the wild type genomic DNA that corresponded to the lec1-1 deletion, a 626 bp ORF was identified. Southern analysis of wild type DNA and the two mutants DNA probed with the short DNA fragment of the ORF revealed that both the wild type and lec1-2 DNA contain the ORF while the lec1-1 genomic DNA did not hybridize. The exact insertion site of the T-DNA in lec1-2 mutant was determined by PCR and sequencing and it was found that the T-DNA was inserted 115 bp upstream of the ORF's translational initiation codon in the 5' region of the gene.

At the site of the T-DNA insertion a small deletion of 21 plant nucleic acids and addition of 20 unknown nucleic acids occurred. These results suggest that in lec1-2 the T-DNA interferes with the regulation of the ORF while in lec1-1 the whole gene is deleted. Thus, both lec1 alleles contain DNA disruptions at the same locus, confirming the identity of the LEC1 locus.

The lec1 Mutants Can Be Complemented by Transformation

To prove that the 7.4 kb genomic wild type fragment indeed contained the ORF of the LEC1 gene, we used a genomic fragment of 3395 bp (SEQ ID NO: 3) within that fragment to transform homozygous lec1-1 and lec1-2 plants. The clone consists of a 3395 bp BstYI restriction fragment containing the gene and the promoter region. The translation start codon (ATG) of the polypeptide is at 1999 and the stop codon is at 2625 (TGA). There are no introns in the gene.

The transformed plants were selected on hygromycin plates and were tested to contain the wild type DNA fragment by PCR analysis. Both transgenic mutants were able to produce viable progeny, that were desiccation tolerant and did not possess trichomes on their cotyledons. We concluded that the 3.4 kb fragment can complement the lec1 mutation and since there is only one ORF in the deleted 2 kb fragment in lec1-1 we suggest that this ORF corresponds to the LEC1 gene.

The LEC1 Gene is a Member of Gene Family

In order to isolate the LEC1 gene two cDNA libraries of young siliques were screened using the 7.4 kb DNA fragment as a probe. Seventeen clones were isolated and after further analysis and partial sequencing they were all found to be identical to the genomic ORF. The cDNA contains 626 bp ORF specifying 208 amino acid protein (SEQ ID NO:1 and SEQ ID NO:2).

The LEC1 cDNA was used to hybridize a DNA gel blot containing Ws-O genomic DNA digested with three different restriction enzymes. Using low stringency hybridization we found that there is at least one more gene. This confirmed our finding of two more Arabidopsis ESTs that show homology to the LEC1 gene.

The LEC1 gene is Embryo Specific

The *lec1* mutants are affected mostly during embryogenesis. Rescued mutants can give rise to homozygous plants that have no obvious abnormalities other than the presence of trichomes on their cotyledons and their production of defective progeny.

Therefore, we expected the LEC1 gene to have a role mainly during embryogenesis and not during vegetative growth. To test this assumption poly (A)+ RNA was isolated from siliques, seedling, roots, leaves, stems and buds of wild type plants and from siliques of *lec1* plants. Only one band was detected on northern blots using either the LEC1 gene as a probe or the 7.4 kb genomic DNA fragment suggesting that there is only one gene in the genomic DNA fragment which is active transcriptionally. The transcript was detected only in siliques containing young and mature embryos and was not detected in seedlings, roots, leaves, stems and buds indicating that the LEC1 gene is indeed embryo specific. In addition, no RNA was detected in siliques of both alleles of *lec1* mutants confirming that this ORF corresponds to the LEC1 gene.

Expression Pattern of the LEC1 Gene

To study how the LEC1 gene specifies cotyledons identity, we analyzed its expression by in situ hybridization. We specifically focused on young developing embryos since the mutants abnormal suspensors phenotype indicates that the LEC1 gene should be active very early during development.

During embryogenesis, the LEC1 transcript was first detected in proglobular embryos. The transcript was found in all cells of the proembryo and was also found in the suspensor and the endosperm. However, from the globular stage and on it accumulates more in the outer layer of the embryo, namely the protoderm and in the outer part of the ground meristem leaving the procambium without a signal. At the torpedo stage the signal was

stronger in the cotyledons and the root meristem, and was more limited to the protoderm layer. At the bent cotyledon stage the signal was present throughout the embryo and at the last stage of development when the embryo is mature and filling the whole seed we could not detect the LEC1 transcript. This might be due to sensitivity limitation and may imply that if the LEC1 transcript is expressed at that stage it is not localized in the mature embryo, but rather spread throughout the embryo.

The LEC1 gene encodes a Homolog of CCAAT binding factor.

Comparison of the deduced amino acid sequence of LEC1 to the GenBank reveals significant similarity to a subunit of a transcription factor, the CCAAT box binding factor (CBF). CBFs are highly conserved family of transcription factors that regulate gene activity in eukaryotic organisms Mantvani, et al., . (1992). Nucl. Acids Res. 20: 1087-1091. They are hetero-oligomeric proteins that consist of between three to four non-homologous subunits. LEC1 was found to have high similarity to CBF-A subunit. This subunit has three domains; A and C which show no conservation between kingdoms and a central domain, B, which is highly conserved evolutionary. Similarly the LEC1 gene is composed of three domains. The LEC1 B domain shares between 75%-85% similarity and 55%-63% identity with different B domains that are found in organisms ranging from yeast to human. Within this central domain, two highly conserved amino acid segments are present. Deletion and mutagenesis analysis in the CBF-A yeast homolog hap3 protein demonstrated that a short region of seven residues (42-48) (LPIANVA) is required for binding the CCAAT box, while the subunit interaction domain lies in the region between residues 69-80 (MQECVSEFISFV) (Xing et al., supra). LEC1 protein shares high homology to those regions.

DISCUSSION

The lec1 mutant belongs to the leafy cotyledon class that interferes mainly with the embryo program and therefore is thought to play a central regulatory role during embryo development. It was shown before that LEC1 gene activity is required to suppress germination during the maturation stage. Therefore, we analyzed the genetic interaction of homozygous double mutants of the different members of the leafy cotyledon class and the abi3 mutant that has an important role during embryo maturation. All the five different combinations of the double mutants showed either an intermediate phenotype or an additive effect. No epistatic relationship among the four genes was found. These findings suggest that the different genes act in parallel genetic pathways. Of special interest was the double mutant lec1/lec2 that was arrested morphologically at the heart stage, but continued to grow

in that shape. This double mutant phenotype indicates that both genes LEC1 and LEC2 are essential for early morphogenesis and their products may interact directly or indirectly in the young developing embryo.

The Role of LEC1 in Embryogenesis

One of the proteins that mediate CCAAT box function, is an heteromeric protein called CBF (also called NFY or CP1). CBF is a transcription activator that regulates constitutively expressed genes, but also participates in differential activation of developmental genes Wingender, E. (1993). Gene Regulation in Eukaryotes (New York: VCH Publishers). In mammalian cells, three subunits have been identified CBF-A, CBF-B and CBF-C and all of which are required for DNA binding. In yeast, the CBF homolog HAP activates the CYC1 and other genes involved in the mitochondrial electron transport Johnson, et al., Proteins. Annu. Rev. Biochem. 58, 799-840. (1989). HAP consists of four subunits hap2, hap3, hap4 and hap5. Only hap2, 3 and 5 are required for DNA binding. CBF-A, B and C show high similarity to the yeast hap3, 2 and 5, respectively. It was also reported that mammalian CBF-A and B can be functionally interchangeable with the corresponding yeast subunits (Sinha et al., supra.).

The LEC1 gene encodes a protein that shows more than 75% similarity to the conserved region of CBF-A. CCAAT motifs are not common in plants' promoters and their role in transcription regulation is not clear. However, maize and Brassica homologs have been identified. A search of the GenBank revealed several Arabidopsis ESTs that show high similarity to CBF-A, B and C. Accession numbers of CBF-A (HAP3) homologs: H37368, H76589; CBF-B (HAP2) homologs: T20769; CBF-C (HAP5) homologs: T43909, T44300. These findings and the pleiotropic affects of LEC1 suggest that LEC1 is a member of a heteromeric complex that functions as a transcription factor.

The model suggests that LEC1 acts as transcription activator to several sets of genes, which keep the embryonic program on and repress the germination process. Defective LEC1 expression partially shuts down the embryonic program and as a result the cotyledons lose their embryonic characteristics and the germination program is active in the embryo.

Example 2

This example demonstrates that LEC1 is sufficient to induce embryonic pathways in transgenic plants.

The phenotype of *lec1* mutants and the gene's expression pattern indicated that LEC1 functions specifically during embryogenesis. A LEC1 cDNA clone under the control of the cauliflower mosaic virus 35S promoter was transferred into *lec1-1* mutant plants in planta using standard methods as described above.

5 Viable dry seeds were obtained from *lec1-1* mutants transformed with the 35S/LEC1 construct. However, the transformation efficiency was only approximately 0.6% of that obtained normally. In several experiments, half the seeds that germinated (12/23) produced seedlings with an abnormal morphology. Unlike wild type seedlings, these 35S/LEC1 seedlings possessed cotyledons that remained fleshy and that failed to expand.

10 Roots often did not extend or extended abnormally and sometimes greened. These seedlings occasionally produced a single pair of organs on the shoot apex at the position normally occupied by leaves. Unlike wild type leaves, these organs did not expand and did not possess trichomes. Morphologically, these leaf-like structures more closely resembled embryonic cotyledons than leaves.

15 The other 35S/LEC1 seeds that remained viable after drying produced plants that grow vegetatively. The majority of these plants (7) flowered and produced 100% *lec1* mutant seeds. Amplification experiments confirmed that the seedlings contained the transgene, suggesting that the 35S/LEC1 gene was inactive in these T2 seeds. No vegetative abnormalities were observed in these plants with the exception that a few displayed defects in apical dominance. A few plants (2) were male sterile and did not produce progeny. One
20 plant that produced progeny segregated 25% mutant *Lec1*⁻ seeds that, when germinated before desiccation and grown to maturity, gave rise to 100% mutant seed, as expected for a single transgene locus. The other 75% of seeds contained embryos with either a wild type phenotype or a phenotype intermediate between *lec1* mutants and wild type. Only 25% of the
25 dry seed from this plant germinated, and all seedlings resembled the embryo-like seedlings described above. Some seedlings continued to grow and displayed a striking phenotype. These 35S/LEC1 plants developed two types of structures on leaves. One type resembled embryonic cotyledons while the other looked like intact torpedo stage embryos. Thus, ectopic expression of LEC1 induces the morphogenesis phase of embryo development in
30 vegetative cells.

Because many 35S/LEC1 seedlings exhibited embryonic characteristics, the seedlings were analyzed for expression of genes specifically active in embryos. Cruciferin A storage protein mRNA accumulated throughout the 35S/LEC1 seedlings, including the leaf-like structures. Proteins with sizes characteristic of 12S storage protein cruciferin

accumulated in these transgenic seedlings. Thus, 35S/LEC1 seedlings displaying an embryo-like phenotype accumulated embryo-specific mRNAs and proteins. LEC1 mRNA accumulated to a high level in these 35S/LEC1 seedlings in a pattern similar to early stage embryos but not in wild type seedlings. LEC1 is therefore sufficient to alter the fate of vegetative cells by inducing embryonic programs of development.

The ability of LEC1 to induce embryonic programs of development in vegetative cells establishes the gene as a central regulator of embryogenesis. LEC1 is sufficient to induce both the seed maturation pathway as indicated by the induction of storage protein genes in the 35S/LEC1 seedlings. The presence of ectopic embryos on leaf surfaces and cotyledons at the position of leaves also shows that LEC1 can activate the embryo morphogenesis pathway. Thus, LEC1 regulates both early and late embryonic processes.

Example 3

This example shows that LEC1 is expressed in zygotes and that the promoters of the invention can therefore be used to target expression in zygotes.

To determine precisely when the LEC1 gene becomes activated, LEC1 RNA levels were analyzed in the egg apparatus of mature female gametophytes before fertilization, in zygotes after fertilization, and in very early stage embryos containing an apical cell and two to three suspensor cells. *In situ* hybridization experiments showed that LEC1 RNA was present in zygotes and early stage embryos but was not detected in female gametophytes. These results show that the LEC1 promoter becomes active in the zygote. The LEC1 is therefore useful to target the expression of sense or antisense versions of regulatory genes or cytotoxic genes to zygotes and early stage embryos.

Example 4

This example shows the identification of a LEC1 homolog from Arabidopsis designated the LEAFY *COTYLEDON1-LIKE* gene.

A Blast search was conducted through the Arabidopsis Database (<http://genome-www.stanford.edu/Arabidopsis/>) using the *LEC1* cDNA nucleotide sequence as a probe to identify homologs of the HAP3 subunit of CCAAT box binding transcription factor from Arabidopsis. The Arabidopsis BAC clone, MNJ7 (Accession Number AB025628), contains a gene, designated *LEC1-Like (LIL)*, that displays the highest amino acid sequence identity with the LEC1 protein of any known Arabidopsis *HAP3* gene. The

nucleotide and amino acid sequences of *LIL* are shown in SEQ ID NO:19 and SEQ ID NO:20, respectively.

The Polymerase Chain Reaction (PCR) was used to amplify the *LIL* gene, which lacks introns. Primers designed to amplify the *LIL* open reading frame contained BamHI and XbaI restriction enzyme sites for cloning purposes. The forward primer, BAMMNJ7-5 sequence is 5'-AGGATCCATGGAACGTGGAGGCTTCCAT-3' with the BamHI site underlined. The reverse primer, 3-MNJ7XBA sequence is 5'-ATCTAGATCAGTACTTATGTTGTTGAGTCG-3' with the XbaI site underlined. The PCR conditions were as follows: 30 cycles of 45 seconds at 94°C, 45 seconds at 53°C, and 3 minutes at 72°C. AmphiTaq DNA polymerase (Perkin Elmer Cetus, 761 Main Ave., Norwalk, CT 06859) was used. PCR products were cloned using the TOPO TA Cloning Kit (Invitrogen, Carlsbad, CA 92008). The nucleotide sequence of the cloned *LIL* gene was determined to confirm its identity.

Accumulation of *LEC1*-LIKE RNA

The *LIL* clone was hybridized with gel blots containing 20 µg of total RNA from leaves, stems, roots, seedlings, and siliques containing either proembryo to heart stage (early) embryos, heart to torpedo stage (middle) embryos, or torpedo to mature (late) embryos. *LIL* RNA was detected only in siliques containing all three stages of embryos. Detection of the *LIL* RNA in siliques from *lec1-1* mutants showed that the RNA detected was not *LEC1*. Thus, like *LEC1*, *LIL* accumulates specifically during embryogenesis.

Complementation of *lec1-1* Mutation by *LIL*

The *LIL* clone was inserted into the *LEC1* promoter/terminator cassette within the plant transformation vector BJ49. The *LEC1* promoter/terminator cassette consists of 1992 bp of DNA 5' of the *LEC1* translation start codon and 770 bp 3' of the *LEC1* cDNA translation stop codon (H.S. Lee, R.W. Kwong, and J.J. Harada, unpublished results). The promoter and terminator are separated by a short polylinker with BglII and AvrII restriction endonuclease sites in which the *LIL* gene was inserted.

This construct was transferred into homozygous *lec1-1* null mutants using in planta transformation procedures with *Agrobacterium tumefaciens* strain GV3101. Unlike *lec1-1* mutant plants whose progeny die following desiccation, plants transformed with the *LIL* construct produced viable seedlings. PCR amplification experiments confirmed that the viable seedlings have the *lec1-1* mutation and the transgene. Seedlings morphologically

resembled wild type rather than *lec1* mutant plants. These results show that the *LIL* gene complements the *lec1* mutation, suggesting overlapping functions for the two genes.

Example 5

This example shows the identification of a *LEC1* ortholog from scarlet runner bean.

Constructing an Embryo-Propagator cDNA Library from the Globular Embryo of Scarlet Runner Bean

A cDNA library was constructed with 150 ng of total RNA isolated from embryo propers (EP) of the scarlet runner bean (SRB; *Phaseolus coccineus*) that were dissected from globular-stage embryo. The SMART PCR cDNA Library Construction Kit (Clontech, cat # K1051-1) was used according to a manufacturer's protocol. Briefly, first strand cDNA was synthesized from EP total RNA using SuperScript II RNase H- reverse transcriptase (Gibco/BRL, cat # 18064-014) in the presence of an Sfi IB-site containing oligo-dT primer (CDSIII/3' PCR primer, Clontech) and a SMART III containing an Sfi IA-site primer (Clontech). Second strand was generated by polymerase chain reaction using 5'- and 3' PCR primers (Clontech). Double-stranded cDNA was digested with Sfi I restriction enzyme (New England BioLabs, cat # 123S) and then size-fractionated over a CHROMA S-400 sepharose column (Clontech). After analyzing collected fractions on a 1.1% agarose gel, four fractions containing high amount of cDNAs in a range of 0.5 kb to 4 kb were pooled and precipitated in an ethanol/salt solution at -20°C overnight. A cDNA pellet was recovered by centrifugation and resuspended in 7 uL of sterile water. cDNA inserts were ligated to Sfi I-digested lambda arms (ITriplEx2, Clontech). Ligation mixtures were packaged into phage heads using Gigapack III Gold Packaging Extract (Stratagene).

Isolation of the Scarlet Runner Bean *LEC1* ortholog cDNA.

The cDNA library was converted from a lambda form to a plasmid form via Cre-Lox system (in vivo excision, Clontech). Colonies were picked randomly for plasmid DNA isolation. The nucleotide sequences of cDNA clones were determined using BigDye terminator, a 5'-TriplEx sequencing primer (Clontech), and the ABI Prism 377 DNA sequencer (Perkin-Elmer Applied Biosystems). The identity of the cDNA clone was determined by BlastX and BlastN analyses.

A BlastX search revealed that a cDNA clone pPCEP112 encoded a protein (SEQ ID NO:22) with high amino acid sequence identity to the Arabidopsis LEC1, especially in the conserved B domain. However, a BlastN result indicated that this SRB cDNA

sequence is more similar to the Arabidopsis *LIL* gene at the nucleotide level. The entire sequence of the pPCEP112 insert was determined to be 988 bp (SEQ ID NO:21).

Spatial Expression Pattern of the *LEC1-Like* Gene in SRB Seeds

To examine the spatial expression pattern of scarlet runner bean *LIL* gene in embryos, we carried out in situ hybridization analyses. The full length cDNA insert of pBSEP112 was used as the template for sense and antisense RNA probe synthesis. The *LIL* mRNA accumulated in both the embryo proper (EP) and suspensor (S) of a 5 days after pollination embryo. In the 7 days after pollination seeds, the RNA is localized intensively in the epidermal layer of the embryo proper and moderately in every cell in both the embryo proper and suspensor. Only background signal was detected using the sense probe. In conclusion, the spatial expression pattern of SRB *LEC1-like* gene in the globular embryo is similar to that of Arabidopsis *LEC1*.

The above examples are provided to illustrate the invention but not to limit its scope. Other variants of the invention will be readily apparent to one of ordinary skill in the art and are encompassed by the appended claims. All publications, databases, Genbank sequences, patents, and patent applications cited herein are hereby incorporated by reference.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Harada, John

Lotan, Tamar

Ohto, Masa-aki

Goldberg, Robert B.

Fischer, Robert L.

Bui, Anhthu

Kwong, Raymond

(ii) TITLE OF INVENTION: Leafy Cotyledon1 Genes and Their Uses

(iii) NUMBER OF SEQUENCES: 18

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Townsend and Townsend and Crew LLP

(B) STREET: Two Embarcadero Center, Eighth Floor

(C) CITY: San Francisco

(D) STATE: California

(E) COUNTRY: USA

(F) ZIP: 94111-3834

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS

(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/804,534

(B) FILING DATE: 21-FEB-1997

(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Bastian, Kevin L.

(B) REGISTRATION NUMBER: 34,774

(C) REFERENCE/DOCKET NUMBER: 023070-077600US

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (415) 576-0200

(B) TELEFAX: (415) 576-0300

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 627 base pairs

007060 : 25051560

(B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS
 (B) LOCATION: 1..627
 (D) OTHER INFORMATION: /product= "LEC1"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATG ACC AGC TCA GTC ATA GTA GCC GGC GCC GGT GAC AAG AAC AAT GGT
 48

Met Thr Ser Ser Val Ile Val Ala Gly Ala Gly Asp Lys Asn Asn Gly
 1 5 10 15

ATC GTG GTC CAG CAG CAA CCA CCA TGT GTG GCT CGT GAG CAA GAC CAA
 96

Ile Val Val Gln Gln Gln Pro Pro Cys Val Ala Arg Glu Gln Asp Gln
 20 25 30

TAC ATG CCA ATC GCA AAC GTC ATA AGA ATC ATG CGT AAA ACC TTA CCG
 144

Tyr Met Pro Ile Ala Asn Val Ile Arg Ile Met Arg Lys Thr Leu Pro
 35 40 45

TCT CAC GCC AAA ATC TCT GAC GAC GCC AAA GAA ACG ATT CAA GAA TGT
 192

Ser His Ala Lys Ile Ser Asp Asp Ala Lys Glu Thr Ile Gln Glu Cys
 50 55 60

GTC TCC GAG TAC ATC AGC TTC GTG ACC GGT GAA GCC AAC GAG CGT TGC
 240

Val Ser Glu Tyr Ile Ser Phe Val Thr Gly Glu Ala Asn Glu Arg Cys
 65 70 75 80

CAA CGT GAG CAA CGT AAG ACC ATA ACT GCT GAA GAT ATC CTT TGG GCT
 288

Gln Arg Glu Gln Arg Lys Thr Ile Thr Ala Glu Asp Ile Leu Trp Ala
 85 90 95

ATG AGC AAG CTT GGG TTC GAT AAC TAC GTG GAC CCC CTC ACC GTG TTC
 336

Met Ser Lys Leu Gly Phe Asp Asn Tyr Val Asp Pro Leu Thr Val Phe
 100 105 110

001020 25091560

ATT AAC CGG TAC CGT GAG ATA GAG ACC GAT CGT GGT TCT GCA CTT AGA
384

Ile Asn Arg Tyr Arg Glu Ile Glu Thr Asp Arg Gly Ser Ala Leu Arg
115 120 125

GGT GAG CCA CCG TCG TTG AGA CAA ACC TAT GGA GGA AAT GGT ATT GGG
432

Gly Glu Pro Pro Ser Leu Arg Gln Thr Tyr Gly Gly Asn Gly Ile Gly
130 135 140

TTT CAC GGC CCA TCT CAT GGC CTA CCT CCT CCG GGT CCT TAT GGT TAT
480

Phe His Gly Pro Ser His Gly Leu Pro Pro Pro Gly Pro Tyr Gly Tyr
145 150 155 160

GGT ATG TTG GAC CAA TCC ATG GTT ATG GGA GGT GGT CGG TAC TAC CAA
528

Gly Met Leu Asp Gln Ser Met Val Met Gly Gly Gly Arg Tyr Tyr Gln
165 170 175

AAC GGG TCG TCG GGT CAA GAT GAA TCC AGT GTT GGT GGT GGC TCT TCG
576

Asn Gly Ser Ser Gly Gln Asp Glu Ser Ser Val Gly Gly Gly Ser Ser
180 185 190

TCT TCC ATT AAC GGA ATG CCG GCT TTT GAC CAT TAT GGT CAG TAT AAG
624

Ser Ser Ile Asn Gly Met Pro Ala Phe Asp His Tyr Gly Gln Tyr Lys
195 200 205

TGA 627

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 208 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Thr Ser Ser Val Ile Val Ala Gly Ala Gly Asp Lys Asn Asn Gly
1 5 10 15

Ile Val Val Gln Gln Gln Pro Pro Cys Val Ala Arg Glu Gln Asp Gln
20 25 30

00160202304000

Tyr Met Pro Ile Ala Asn Val Ile Arg Ile Met Arg Lys Thr Leu Pro
35 40 45

Ser His Ala Lys Ile Ser Asp Asp Ala Lys Glu Thr Ile Gln Glu Cys
50 55 60

Val Ser Glu Tyr Ile Ser Phe Val Thr Gly Glu Ala Asn Glu Arg Cys
65 70 75 80

Gln Arg Glu Gln Arg Lys Thr Ile Thr Ala Glu Asp Ile Leu Trp Ala
85 90 95

Met Ser Lys Leu Gly Phe Asp Asn Tyr Val Asp Pro Leu Thr Val Phe
100 105 110

Ile Asn Arg Tyr Arg Glu Ile Glu Thr Asp Arg Gly Ser Ala Leu Arg
115 120 125

Gly Glu Pro Pro Ser Leu Arg Gln Thr Tyr Gly Gly Asn Gly Ile Gly
130 135 140

Phe His Gly Pro Ser His Gly Leu Pro Pro Pro Gly Pro Tyr Gly Tyr
145 150 155 160

Gly Met Leu Asp Gln Ser Met Val Met Gly Gly Gly Arg Tyr Tyr Gln
165 170 175

Asn Gly Ser Ser Gly Gln Asp Glu Ser Ser Val Gly Gly Gly Ser Ser
180 185 190

Ser Ser Ile Asn Gly Met Pro Ala Phe Asp His Tyr Gly Gln Tyr Lys
195 200 205

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3395 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

AGATCCAAAA CAGGTCATGG ACTGGGCCGT AAACCTCTATC CAAAATTCTT CATGTTTTTC 60
CATCTTTCAA AAATCTTTAT CCACCATTCC ATTACTAGGG TGTTGGTTTT ATTTTATTTG 120
TTGATTAATT ATGTATTAGA AAATGTAAAG CAATATTCAA TTGTAACATG CATCATCTAA 180

| | |
|---|------|
| CACCAATATC TTGTAATAAC CTTTGTAAAT TTTCTATAA ACATTTTAAA AGGCTAATTT | 240 |
| AAATAAAAAT TACAATAAAC GTGATAACTC ACTTTCGTAA CGCATATTTA TTCAAATATA | 300 |
| CCAAAATTTA CCATTTTAAG TAAGAGAATC TTTTAAAAAT TAATTTTCAA TTTCATTAAT | 360 |
| TAAGAAACAA AGAATTTACT GAAACCTATA TTTTATTAAA TTTTAATAAA ATATATGACT | 420 |
| AAAATAACGT CACGTGAATC TTTCTCAGCC GTTCGATAAT CGAATACTTT ATTGACTAAG | 480 |
| TATTTATTTA GAAAATTTA AACAACTT AATTTCTAGA AACAAAGAGA GCCTCATATG | 540 |
| TATAAAAATC TTCTTCTTAT CTTTCTTTCT TTCTTAATAG TCTTTATTTT TACTTAATTA | 600 |
| CTTTGGTAAT TTGTGAAAAA CACAACCAAT GAGAGAAGAG CAGTTTGACT GGCCACATAG | 660 |
| CCAATGAGAC AAGCCAATGG GAAAGAGATA TAGAGACCTC GTAAGAACCG CTCCTTTGCC | 720 |
| ATTTGTATCA TCTCTCTATA AAACCACTCA ACCATCAACC TNTCTTTGCA TGCAACAAAT | 780 |
| CACTCAAATA ATTATTTTAT AAAGAACAAA AAAAAAAGA CGGCAGAGAA ACAATGGAAC | 840 |
| GTGGAGCTCC CTTCTCTCAC TATCAGCTAC CCAAATCCAT CTCTGGTAAT CTAAGTGGCT | 900 |
| ATTTGTATAC AGTATATACT TGCCTCCATG TATATTTATA TTCTCGTGAA AAATTGGAGA | 960 |
| CATGCTTTAT GAATTTTATG AGACTTTGCA ACAACGAACG AGATGCTTTC TCTCTAGAAA | 1020 |
| TTTAAATTTA GATTTGTGAA GGTTTTGGGA ATGGCCCGGA GAAGACGATT TTATATATAC | 1080 |
| ATGCATGCAA GAGTTTGATA TGTATATTGT TTCATCATGG CTGAGTCAAA GTTTTATCCA | 1140 |
| AATATTTCCA TGGTGTGGTA TTAGTTAAAC AAATCTCTCG TATGTGTCAT TGAATATACC | 1200 |
| CGTGCATGTA CCAGGAATGT TTTTGATTCT AAAAACGTTT TTTTCTTTGT TGTAACGGTT | 1260 |
| GAGTTTTTTT CTTCGTTTCA AAACGAGATT CTCGTTTGTC TCTTCCCTTG TCTAAAAACA | 1320 |
| TCTACGGTTC ATGTGATTCA AAAACACTAA AAAAATATAA ACTCATTTTT TTTTAATACT | 1380 |
| TAACATTTAA ACTATATATA TATATATATA TATATATATC TTATACTAGT CCCAAGTTTT | 1440 |
| AGTGTGAGGT TTTTTTATTC AAAATCTATC AGTACATTTT TTGGAAAAGA ACTAAGTGAA | 1500 |
| ATTTTCTCCA AATTTTCCTT TTAATTGGA TTTTAAATT ACTGGATGTC ATTAACTTTA | 1560 |
| ATCTTTTGAT TCTTTCAACG TTTACCATTG GGAACCTTCA CATGAAATAA ATGTCTACTT | 1620 |
| TATTGAGTCA TACCTTCGTC AACATAAATT AATTGATGTT CTTCTCCAAA TTTTGAGTTT | 1680 |
| TTGGTTTTTC TAATAATCTT AACGAAAGCT TTTTGGTATA CATGTAAAAC GTAACGGCAA | 1740 |
| GAATCTGAAC AGTCTACTCA ACGGGGTCCA TAAGTCTAGA ATGTAGACCC CACAACTTA | 1800 |
| CTCTTATCTT ATTGGTCCGT AACTAAGAAC GTGTCCCTCT GATTCTCTTG TTTTCTTCTA | 1860 |
| ATTAATTCGT ATCCTACAAA TTTAATTATC ATTTCTACTT CAACTAATCT TTTTTTATTT | 1920 |
| CCTAAAGATT TCAATTTCTC TCTGTATTTT CTATGAACAG AATTGAACTT GGACCAGCAC | 1980 |

AGCAACAACC CAACCCCAAT GACCAGCTCA GTCATAGTAG CCGGCGCCGG TGACAAGAAC2040
 AATGGTATCG TGGTCCAGCA GCAACCACCA TGTGTGGCTC GTGAGCAAGA CCAATACATG 2100
 CCAATCGCAA ACGTCATAAG AATCATGCGT AAAACCTTAC CGTCTCACGC CAAAATCTCT 2160
 GACGACGCCA AAGAAACGAT TCAAGAATGT GTCTCCGAGT ACATCAGCTT CGTGACCGGT 2220
 GAAGCCAACG AGCGTTGCCA ACGTGAGCAA CGTAAGACCA TAACTGCTGA AGATATCCTT 2280
 TGGGCTATGA GCAAGCTTGG GTTCGATAAC TACGTGGACC CCCTCACCGT GTTCATTAAC 2340
 CCGTACCGTG AGATAGAGAC CGATCGTGGT TCTGCACTTA GAGGTGAGCC ACCGTCGTTG 2400
 AGACAAACCT ATGGAGGAAA TGGTATTGGG TTCACGGCC CATCTCATGG CCTACCTCCT 2460
 CCGGGTCCTT ATGGTTATGG TATGTTGGAC CAATCCATGG TTATGGGAGG TGGTCGGTAC 2520
 TACCAAAACG GGTCGTCGGG TCAAGATGAA TCCAGTGTTG GTGGTGGCTC TTCGTCTTCC 2580
 ATTAACGGAA TGCCGGCTTT TGACCATTAT GGTCAGTATA AGTGAAGAAG GAGTTATTCT 2640
 TCATTTTTAT ATCTATTCAA AACATGTGTT TCGATAGATA TTTATTTTT ATGTCCTTATC 2700
 AATAACATTT CTATATAATG TTGCTTCTTT AAGGAAAAGT GTTGTATGTC AATACTTTAT 2760
 GAGAAACTGA TTTATATATG CAAATGATTG AATCCAAACT GTTTTGTGGA TTAAACTCTA 2820
 TGCAACATTA TATATTTACA TGATCTAAAG GTTTTGTAAT TCAAAAGCTG TCATAGTTAG 2880
 AAGATAACTA AACATTGTAG TAACCAAGTT TAATTTACTT TTTGAGTTT ACATAACTAA 2940
 CCAAGCCAAA AGGTTATAAA ATCTAAATTC GTTGAGTTGT CAAACTTCTG AAGATTGCTA 3000
 TCCTCTTTGA GTTGCTTTCT TTTGGGTGCT TGAGTTTCAT TAGGCTGAGC TGA CTCTCTGTTT 3060
 CTCTCTAGTC TTTCATCTCT GTCTTTTCCA AGGATTCATA ACGTTGGTCG CTCTCTGTTT 3120
 CTGCCTACAC TTCTTCAAGG GATCATTACT GAGGCTAAGA GTTAAAGACC TGAACCATGG 3180
 TTTTCTGTAA CTGGTTCAAG TTCATTCTCC GGTATTGTG TGGTTATCTT TCGGTTAGAT 3240
 TGAAACCCAT ATGTTTGCTC TGTTTCTTCT AGTTCCAAGT TTAATTTCCG GTTATTGTTT 3300
 GGCTTTTTTAA AAGTTTTTAA GGTCTATTCT ATGTAAAGAC TATTCTACGT ACGTACATTT 3360
 ATCGCAAAAT TGAAAGATTA TAAAAAAAAT TGAAA 3395

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7560 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AATTNACCCT CACTAAAGGG AACAAAAGCT GGGTACCGGG CCCCCCCTCG
AGGTCGACGG 60

TATCGATAAG CTTGATATCG AATTCGTGGC CATTAGACCC ATAACTATAT
GACGATGTTA 120

AAGAGAAAAT AAATCATAAA TAAAATAAGA GTCCTTATCA ATAAACCTAA
TTGGCTAATT 180

TCAACCTCAA AGAGTAGTAG GAACAGGTAA GGTGAAGCCA AACAGCTCCT
TTTACAGTTG 240

GACCACTAGA GCTGATCTGG CATACAAAGT ATGCTTATTG GGCTGTCACG
GCCCATCCGC 300

AAAATGTCGT TGGTTACGAA GCATCCACGA CATAGACGGT GCCACATGTT
AGAAAAGTGT 360

TTCGGCGATC AAGATTGTGT CCACATCATT AGACGTCTGA ACTGTCCACG
TGTCTATCAA 420

AGCTGGCGTC AAACATTACG TTTTCGTCGT TTGCGCCTCC TAGTTCACAC
GTGCAACGAA 480

CGCGTGCGAC GTATCAAAAT TGTTAATTTT AGCCATGTAT AAAGAATATC
TACAAAATTA 540

ACCTCAGGAA TATTTTTGTT TTTTCAATTG AGGCCATAAT ATACNTNCCG
ATNGAAAAAT 600

TTTNCANCAT ATCNCTAATA TCAAAAATT ATGATGTTAG TAAACGTAAA
AAATTTACAC 660

AAAATAANTT TCACAAAAC TANNGGGGAA ATTGGAACAA ANAAAAGACT
GGTGAGTGAT 720

AAGCGATGAT GGCCGGTGAA TCAGGTAGCC GTCCTACAAC GTGGTTGATT
TTGAGCAAAC 780

TCCTATCTAC TCTTCACACT ATTGGAAATC CAAAATGTC GTCACACCAT
AATAATGTGA 840

ATTTTGTTAT GGAATTTGAG GGAAACAGTA GATATATGTT TCAACCAGTG
AAAGTTACCC 900

007020 23051560

TCCTTTGGAC ATATCTACGA NAGTAGAAAG TAGAAACATT CACTAAACGT
GACAACTTTA 960

TAAATTTTCT TTTTGTAAC TTTCTTTAGA TTTATTTACG ANAAGAGAAA
TATAAACGTC 1020

ATGCTAATAA AAAATGCATT ATTTTCTACC ATCTAGCTAG AATATTGATC
AAGTCTTCAC 1080

GTTTTTTGTT TATCTCTTCT CTCATAGGCA TGTCCACAAA AGGGTAAGTT
TTACTGGTTC 1140

AAAATATTGC ATGAGTACTA CTAAGCTCGT ATAGTTTGAT CTTACTATCA
TTGCGATGAG 1200

GGTTGTTAGT TTGGAAGAAA TAAGGATTTA TGCAAATGGT AATCATTATG
TCTGCTATTT 1260

AAGAAGTAAA TTATGATGCT TGTTGCGTGA ACATATTAAA TTTGCGAAAA
ATAAGCAAGG 1320

ATACACGAGA GAAGCTCAGA TATTCACGTA ACGATGTTTC ATCTCTTCTC
ATTGAGGAAA 1380

CATATGGCCA TGATATAGCT AATAAGCCTA CGGGATTGTC NTTTCAACGC
CGAATCTACC 1440

AAACTGTTCC ATCTCTTATT ATATATAGTT TGGTTATTTA AGTAATTAGA
TGCATCATAA 1500

TCTTTTTTTC TGCCAGTTGT AATGCAGATA AAAATATATT GGTGTTCTA
AGGATTGTTT 1560

AAACGTGCAT GTGTACAAGT TATTATTTAT ATACTTTCAT CTACATGCGA
TGC GTTATTT 1620

ATAATGATAA AACTAAGATT TTTAGTTAAA TTTAATAAAG AGCTTACGAG
CTACAATTAA 1680

TTAGAAATGG TTGCTCAGAA ATCAGAATAC TATATATGAA AAAAGAAGTT
GGTATACTTG 1740

AAAAAAGAAA AAAC TACTTG AAAAGATGGT AAAAGATATA GAACGAGTAT
ATATCTTACT 1800

CAAGCACGAT AGAAGTTTGT ATCAAAACAT TGC GTTCCAA ACCAATGTTT
GAAGATGGTC 1860

AAAGGTGCTA CTCATGATGT GGTGCGAAGA AGCTTACGAA AAATTCTGCA
ATGAGAGATA 1920

001020:25051560

ACTTTATGGG CTGCTTGTTT AATATATTGA AAATCATGGT AGACAACACC
AAACTCTCCT 1980

TTACCAGAAG TCATATTTCC TTAACCTCAG AATAAGTAAA TCTTCTAGTT
TATTATTTGA 2040

AAGTTGAGCG TATAATTGCA ATGAACTTT TACCAATTCA CCGCCTCCTA
ACTGAGTTGT 2100

TGTATTATCC TATCTCTTTA GCTATCCTTT CCTTGCTCTT GCTCCACCTG
CATGTGGCCT 2160

CTTTATTTAT AATCTCTCTA GATTCTGCTA AAGATGTNTG TTCAAAATGG
TTTATCTTTA 2220

AGGGAAGCAA AGTGAATGGA AACATTTAAA GAAAAAAAAA ACTTTTAGCA
GAGTTCCATG 2280

AGATTTTATA CTGATGATAA CTAAAATAAT CTTATATGCG TAAGATTATT
TTAGTTCTAA 2340

ACTTCATTTT GAAATGAGAG GTCATTGGCC AGGAAAGATT CAATATTGGT
TCTTTGTTAA 2400

TTCTCGTTGG TTTGTTTTTA GTATGGGCTA GATCCAAAAC AGGTCATGGA
CTGGGCCGTA 2460

AACTCTATCC AAAATTCTTC ATGTTTTTCC ATCTTTCAA AATCTTTATC
CACCATTCCA 2520

TTACTAGGGT GTTGGTTTTA TTTTATTTGT TGATTAATTA TGTATTAGAA
AATGTAAAGC 2580

AATATTCAAT TGTAACATGC ATCATCTAAC ACCAATATCT TGTACTAACC
TTTTGTAATT 2640

TTCCTATAAA CATTTTAAAA GGCTAATTTA AATAAAAATT ACAATAAACG
TGATAACTCA 2700

CTTTCGTAAC GCATATTTAT TCAAATATAC CAAAATTTAC CATTTTAAGT
AAGAGAATCT 2760

TTTTAAAATT AATTTTCAAT TTCATTAATT AAGAAACAAA GAATTTACTG
AAACCTATAT 2820

TTTATTAAAT TTTAATAAAA TATATGACTA AAATAACGTC ACGTGAATCT
TTCTCAGCCG 2880

TTCGATAATC GAATACTTTA TTGACTAAGT ATTTATTTAG AAAATTTTAA
ACAACACTTA 2940

ATTTCTAGAA ACAAAGAGAG CCTCATATGT ATAAAAATCT TCTTCTTATC
TTTCTTTCTT 3000

TCTTAATAGT CTTTATTTTT ACTTAATTAC TTTGGTAATT TGTGAAAAAC
ACAACCAATG 3060

AGAGAAGAGC AGTTTGACTG GCCACATAGC CAATGAGACA AGCCAATGGG
AAAGAGATAT 3120

AGAGACCTCG TAAGAACCGC TCCTTTGCCA TTTGTATCAT CTCTCTATAA
AACCACTCAA 3180

CCATCAACCT NTCTTTGCAT GCAACAAATC ACTCAAATAA TTATTTTATA
AAGAACAAAA 3240

AAAAAAAGAC GGCAGAGAAA CAATGGAACG TGGAGCTCCC TTCTCTCACT
ATCAGCTACC 3300

CAAATCCATC TCTGGTAATC TAAGTGGCTA TTTGTATACA GTATATACTT
GCCTCCATGT 3360

ATATTTATAT TCTCGTGAAA AATTGGAGAC ATGCTTTATG AATTTTATGA
GACTTTGCAA 3420

CAACGAACGA GATGCTTTCT CTCTAGAAAT TTAAATTTAG ATTTGTGAAG
GTTTTGGGAA 3480

TGGCCCGGAG AAGACGATTT TATATATACA TGCATGCAAG AGTTTGATAT
GTATATTGTT 3540

TCATCATGGC TGAGTCAAAG TTTTATCCAA ATATTTCCAT GGTGTGGTAT
TAGTTAAACA 3600

AATCTCTCGT ATGTGTCATT GAATATACCC GTGCATGTAC CAGGAATGTT
TTTGATTCTA 3660

AAAACGTTTT TTTCTTTGTT GTAACGGTTG AGTTTTTTTC TTCGTTTCAA
AACGAGATTC 3720

TCGTTTGTCT CTTCCCTTGT CTAAAAACAT CTACGGTTCA TGTGATTCAA
AAACACTAAA 3780

AAAATATAAA CTCATTTTTT TTTAATACTT AACATTTAAA CTATATATAT
ATATATATAT 3840

ATATATATCT TATACTAGTC CCAAGTTTTA GTGTGAGGTT TTTTATTCA
AAATCTATCA 3900

GTACATTTTT TGGAAAAGAA CTAAGTGAAA TTTTCTCCAA ATTTTCCTTT
TACTATTGAT 3960

TTTTTAATTA CTGGATGTCA TTAAC TTAA TCTTTTGATT CTTTCAACGT
TTACCATTGG 4020

GAACCTTCAC ATGAAATAAA TGTCTACTTT ATTGAGTCAT ACCTTCGTCA
ACATAAATTA 4080

ATTGATGTTT TTTCTCCAAAT TTTGAGTTTT TGGTTTTTCT AATAATCTTA
ACGAAAGCTT 4140

TTTGGTATAC ATGTAAAACG TAACGGCAAG AATCTGAACA GTCTACTCAA
CGGGGTCCAT 4200

AAGTCTAGAA TGTAGACCCC ACAAACTTAC TCTTATCTTA TTGGTCCGTA
ACTAAGAACG 4260

TGTCCCTCTG ATTCTCTTGT TTTCTTCTAA TTAATTCGTA TCCTACAAAT
TTAATTATCA 4320

TTTCTACTTC AACTAATCTT TTTTATTTC CTAAAGATT CAATTTCTCT
CTGTATTTTC 4380

TATGAACAGA ATTGAACTTG GACCAGCACA GCAACAACCC AACCCCAATG
ACCAGCTCAG 4440

TCATAGTAGC CGGCGCCGGT GACAAGAACA ATGGTATCGT GGTCCAGCAG
CAACCACCAT 4500

GTGTGGCTCG TGAGCAAGAC CAATACATGC CAATCGCAAA CGTCATAAGA
ATCATGCGTA 4560

AAACCTTACC GTCTCACGCC AAAATCTCTG ACGACGCCAA AGAAACGATT
CAAGAATGTG 4620

TCTCCGAGTA CATCAGCTTC GTGACCGGTG AAGCCAACGA GCGTTGCCAA
CGTGAGCAAC 4680

GTAAGACCAT AACTGCTGAA GATATCCTTT GGGCTATGAG CAAGCTTGGG
TTCGATAACT 4740

ACGTGGACCC CCTCACCGTG TTCATTAACC GGTACCGTGA GATAGAGACC
GATCGTGGTT 4800

CTGCACTTAG AGGTGAGCCA CCGTCGTTGA GACAAACCTA TGGAGGAAAT
GGTATTGGGT 4860

001000 25097500

TTCACGGCCC ATCTCATGGC CTACCTCCTC CGGGTCCTTA TGGTTATGGT
ATGTTGGACC 4920

AATCCATGGT TATGGGAGGT GGTCGGTACT ACCAAAACGG GTCGTCGGGT
CAAGATGAAT 4980

CCAGTGTGG TGGTGGCTCT TCGTCTTCCA TTAACGGAAT GCCGGCTTTT
GACCATTATG 5040

GTCAGTATAA GTGAAGAAGG AGTTATTCTT CATTTTTATA TCTATTCAAA
ACATGTGTTT 5100

CGATAGATAT TTTATTTTTA TGTCTTATCA ATAACATTC TATATAATGT
TGCTTCTTA 5160

AGGAAAAGTG TTGTATGTCA ATACTTTATG AGAACTGAT TTATATATGC
AAATGATTGA 5220

ATCCAACTG TTTTGTGGAT TAACTCTAT GCAACATTAT ATATTACAT
GATCTAAAGG 5280

TTTTGTAATT CAAAAGCTGT CATAGTTAGA AGATAACTAA ACATTGTAGT
AACCAAGTTT 5340

AATTTACTTT TTTGAGTTTA CATAACTAAC CAAGCCAAAA GGTATAAAAA
TCTAAATTCG 5400

TTGAGTTGTC AAACCTCTGA AGATTGCTAT CCTCTTTGAG TTGCTTTCTT
TTGGGTGCTT 5460

GAGTTTCATT AGGCTGAGCT GACTCGTTGC TCTCTAGTCT TTCATCTCTG
TCTTTTCCAA 5520

GGATTCATAA CGTTGGTCGC TCTCTGTTTC TGCCTACACT TCTTCAAGGG
ATCATTACTG 5580

AGGCTAAGAG TTAAAGACCT GAACCATGGT TTTCTGTAAC TGGTTCAAGT
TCATTCTCCG 5640

GTTATTGTGT GGTTATCTTT CGGTTAGATT GAAACCCATA TGTTTGCTCT
GTTTCTTCTA 5700

GTTCCAAGTT TAATTTCCGG TTATTGTTTG GCTTTTTAAA AGTTTTTAAG
GTCTATTCTA 5760

TGTAAAGACT ATTCTACGTA CGTACATTTA TCGCAAAATT GAAAGATTAT
AAAAAAAATT 5820

GAAAGATCCA AAGGAAACCA ATAGATTAAA CTAAAATGTA GTATCCTTTT
TATCATTTTA 5880

001020 25051560

GGCTATGTTT TCTTTTAAGA AAGCTTTGGT AGTTAACTCT GTTTAAAAGA
AAAAAAAGAG 5940

ATGCATAAAT TAAATTTAAG TTTCTAGAAC TTTTGGATAA ACATATTAAG
CTAAAGAAAT 6000

TAAACTAAAG GGC GTAAATG CAAGCTTGTT ATGCGTTATT GAAAACATTA
CCTCTAAATT 6060

AAATAGCCCA ATATTGAAAA CCTTAAGCTT CTTTGATCCC CTTAACTTGT
TTGTCCACCA 6120

AGTATTAGTT CATCTCTTAA CACGGCAACT CGAAACGGCA CAATGGACAA
ACATGGTCTT 6180

TCAAAAACCA CTTCCCAATA CATCCATCGT CAAACTCGTG GCCACATGGT
AAGGTCACCA 6240

CTATTTCTCC CTTTTCAAAC TCCTCCAAAC AAATTGTGCA CACACTGGCG
TCAGAGTTGG 6300

ATTTCTTCTT ATTATTATAT ACTTTCCTTG CCAAACGGTC AACCACAAAC
TTATTTGCCG 6360

GTCTAATTAA CTCGATATTA TTGGTGGTCT CATCAAACGA GTCAATCCGA
GGAGGAGGTG 6420

GAACAATGAC TTTACAGTAC ATGTAACTA ACGTAGCACA AACTGAAGAG
TCTACCATAG 6480

AAATCGACTT ACAGATTCGT TCAGTGAGTT GAGAGTTAGC AATGTCAACA
TATTGTTCGG 6540

AGAGCCCTGC TGAGTACAAC CATTCATTCA GTTTTTTCGA GTCATTAGGG
TAGGAGGATA 6600

TGACACCTTC GTAGTCATTG TACGAGAGAA CGAAATTGG TGGAAGACTA
ATTGATGTGT 6660

CCGATCTTCG GGCACCTACG CAGATTTTGA ATGATCCAGC ATCTTGTGAT
TTCGGTTTGA 6720

GGTCTATTTT GCCGCCAAAG GATATTTCCG CTTCCATAGC TATCAAAGAG
AAAGAAAAAT 6780

AGTGAATCCA AGGTTTAGGG TTTCTTTTCT TTGTCTTNCT TATATATAGA
GGCGCTAGAT 6840

TGTATTAAGG ATTATACATA TATATAAGTA ATTGCAATTT GTGAGTTTAT
CCTTATTCAT 6900

TTTTAATTTT ATTTACCTTT ATTTAGTTGA TATTGTGTCC TTTTCCTAGG
TAGCATTTC 6960

TTCCATCTGT GTTAATTATT AGCATTTCCT TTCCTTTGTC TTATTTCCT
TTATTTCGTA 7020

GGAAGAAATC CTTTATGNAC CCCATCTTGG CTGAGAACTT GAGATGATTT
TAAATCCTCA 7080

AAAATTATTC AATTTATGAT TTCGAAATTG ATATACACTT TATATTTTCT
CCTAAAAAAC 7140

CATATTGTAC TAAGAAAAGT AGAAAACCAG ACTTTTAAAT ATGTTAGATT
TTAATTGGGT 7200

TCTTAAAGTG TTTTAGCGTT TNACACCGGT TATTCTCAA AATCCAACT
CTATAATTAT 7260

AGTTTTTAAG TATAAATTAA TCCGGTTGGC CCAATTAGTG GACCGTTTAA
AGAGTAGACA 7320

CTTTTTTTTT TATATATCGA CTACCATAAA ACTTTAACGA TTAATATTTT
TGGATAATAA 7380

GCGATCGTTT TGAGGCGTCC CAATTTTTTT TGTTTCTTTT TATATGAGAA
ATGGGTTTAA 7440

GAAAAACTGC AATTTTGTCC ATAAAGCTAG TCAGAATTCC TGCAGCCCGG
GGGATCCACT 7500

AGTTCTAGAG CGGCCGCCAC CGCGGTGGAG CTCCAATTCG CCCTATAGTG
AGTCGTATTA 7560

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

0016020125097560

Met Pro Ile Ala Asn Val Ile

1 5

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 12 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Ile Gln Glu Cys Val Ser Glu Tyr Ile Ser Phe Val

1 5 10

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GGAATTCAGC AACAACCCAA CCCCA

25

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GCTCTAGACA TACAACACTT TTCCTTA

27

(i) SEQUENCE CHARACTERISTICS:

- (ii) MOLECULE TYPE: DNA

ATGACCAGCT CAGTCATAGT AGC

23

(i) SEQUENCE CHARACTERISTICS:

- (ii) MOLECULE TYPE: DNA

GCCACACATG GTGGTTGCTG CTG

23

(i) SEQUENCE CHARACTERISTICS:

- (ii) MOLECULE TYPE: DNA

GAGATAGAGA CCGATCGTGG TTC

23

(2) INFORMATION FOR SEQ ID NO:12:

- (A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TCACTTATAC TGACCATAAT GGTC

24

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GCATAGATGC ACTCGAAATC AGCC

24

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GCTTGGTAAT AATTGTCATT AG

22

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (ii) MOLECULE TYPE: DNA

CTAAAAACAT CTACGGTTCA 20

(i) SEQUENCE CHARACTERISTICS:

- (ii) MOLECULE TYPE: DNA

TTTGTGGTTG ACCGTTTGGC 20

(i) SEQUENCE CHARACTERISTICS:

- (ii) MOLECULE TYPE: peptide

Leu Pro Ile Ala Asn Val Ala
1 5

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Met Gln Glu Cys Val Ser Glu Phe Ile Ser Phe Val
 1 5 10

SEQ ID NO:19: *Arabidopsis* L1L gene (pMNJ7 sequence)

ATGGCAGAGG GCAGTATGCG TCCTCCAGAA TTCAACCAGC CTAACAAAAC
 CAGTAATGGT GGTGAGGAGG AGTGACGGT GAGGGAGCAA GACAGGTTCA
 TGCCTATTGC CAACGTGATA CGGATCATGC GGAGGATCTT ACCTGCTCAC
 GCCAAGATCT CAGATGACTC CAAGGAGACG ATCCAAGAGT GTGTTTCGGA
 GTACATCAGC TTCATAACAG GGGAGGCTAA TGAGCGGTGC CAGCGGGAAC
 AGCGCAAGAC CATCACTGCT GAGGACGTCT TGTGGGCAAT GAGCAAGCTC
 GGTTTTGATG ACTACATCGA ACCCCTCACG TTGTACCTCC ACCGCTACAG
 AGAGTTGGAA GGTGAAAGAG GGGTTAGCTG CAGTGCTGGG TCCGTTAGTA
 TGACCAACGG CTTGGTGGTC AAGAGGCCTA ATGGGACCAT GACCGAGTAT
 GGAGCCTACG GGCCTGTGCC AGGGATTCAC ATGGCGCAGT ACCATTATCG
 TCATCAGAAC GGGTTTGTTT TCAGTGGTAA CGAACCTAAT TCTAAGATGA
 GTGGTTCATC TTCAGGAGCA AGTGGCGCCA GAGTTGAAGT ATTTCCGACT
 CAACAACATA AGTACTGA

SEQ ID NO:20: *Arabidopsis* L1L protein

MAEGSMRPPE FNQPNKTSNG GEEECTVREQ DRFMPIANVI RIMRRILPAH
 AKISDDSKET IQECVSEYIS FITGEANERC QREQRKTITA EDVLWAMSKL
 GFDDYIEPLT LYLHRYRELE GERGVSCSAG SVSMTNGLVV KRPNGTMTEY
 GAYGPVPGIH MAQYHYRHQN GFVFSGNEPN SKMSGSSSGA SGARVEVFPT
 QQHKY

SEQ ID NO:21: *Phaseolus* gene

GATCTCTCAACCCAACCCTTTTCATTTTCATTTTCATTTTCATTTTCCATCACTTCACTGTC
 ACCATGGAAAG
 TGGAGGCTTTTCATGGCTACCGCAAGCTCCCCAACACCACCTCTCCTGGGTTGAAGCTGTCAG
 TGTCAGACATG
 AACAACGTGAACACGAGTAGGCAGGTAGCAGGAGACAACAACCACACAGCGGATGAGAGCAA
 CGAATGCACTG
 TGAGGGAGCAAGACCGTTTCATGCCAATTGCAAATGTGATCAGGATCATGCGAAAGATTCTT
 CCTCCACATGC
 CAAGATCTCAGGTGATGCCAAAGAAACAATTCAAGAGTGTGTGTCTGAGTACATCAGCTTTA
 TCACCGGAGAG
 GCAAACGAGCGTTGCCAGAGGGAACAACGCAAGACCATAACTGCTGAGGACGTGCTTTGGGC
 CATGAGCAAGC

TTGGATTTGATGATTACATGGAGCCACTGACCATGTACCTTCACAGGTATCGTGAGCTTGAG
 GGTGACCGAAC
 CTCCATGAGAGGTGAATCATTGGGGAAGAGGACTATTGAATACGCCCCCTATGGGTGTTGGCG
 TTGCTACTGCT
 TTTGTGCCACCACAGTTTCACCCAAATGGATACTATGGTCCTGCCATGGGAGCTTACGTTGC
 GCCACCAAATG
 CTGCGTCCTCTCATCACCATGGAATGCCAAATACTGAACCGAATGCTCGCTCCATGTGAATT
 GATGATGATGA
 GGAGGAGGAGGAGGAAGACGACGAGTGTTGAGTTAGTAGAAGAAGAATACTTTAATTAATTA
 GCTTAACTCTC
 GGTAATTAGAGTACTGTTGTTGAGGGTACGTAGTAACTTTATAATTAAGGGGATGGATGGG
 ATTAAGGAGTT
 CTGATATTCCTAATCCTAATCAGGCCTATGTTAATTTATGTAATAACTCTGCTTATGTTTTT
 GGATTTTCTGA
 TGTGTTCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

SEQ ID NO:22: *Phaseolus* protein

MESGGFHYRKLPLNTTSPGLKLSVSDMNNVNTSRQVAGDNNHTADESNECTVREQDRFMPIA
 NVIRIMRKILP
 PHAKISGDAKETIQECVSEYISFITGEANERCQREQRKTITAEDVLWAMSKLGFDDYMEPLT
 MYLHRYRELEG
 DRTSMRGESLGKRTIEYAPMGVGVATAFVPPQFHPNGYYGPAMGAYVAPPNAASSHHHGMPN
 TEPNARSM

SEQ ID NO:23: 5' untranslated region

| | | | | | |
|-------------|------------|------------|-------------|-------------|------------|
| atgattctct | tcctcctctt | caaatggagt | ttcaagctcg | tgggttttca | aaggaagagg |
| gtctctctct | caggatataa | tctcaccatt | aaaaatgtga | aaatcgcatc | tcttgggatg |
| tctgttactg | tgaaaagttg | ttactttttt | tctgtattat | gctttttgtt | caactttgga |
| acgtttatta | gaagcttggt | tggtagagac | ctcctaaaca | taagagtcta | attttttttc |
| atgtgagctt | tgcggtatca | tttgattcta | gattgggtga | cattctcttc | ctcttgatat |
| ctcagcttaa | agcattaaac | tttgagata | tcaatcagat | ctggtgcatc | actgaacact |
| cttttacagt | gttggtttat | accacttcta | agcagtgttt | tggtgtgccg | tcattacaag |
| ttttggatta | ttagtcttta | gatagtgtaa | ccatggtgga | gtctatatat | tctgcggaac |
| tactttccaa | tttttgattt | tgagctcctt | ctggtgatag | agctttgagt | ttttgataag |
| cagacgttcg | caagacogtc | acgggtaaaa | gaaagcggga | cagcgatagt | gactcatctc |
| atgtcttgga | gaagcttggt | gggagtatga | tgaagcggca | aacaagggtg | aagctggagc |
| tgattaatgt | gatggagaag | atggaagtgg | agagaatacg | ggagaagatg | cataatcagt |
| aacaggaaac | cgagaggatg | acacagaatg | aagaagcacg | ccgtgaggaa | gcttggaggc |
| acttgtctct | catctctttc | atcagaagtg | ttactgggtg | gaagcaagag | atggcacgca |
| agtggtgaatt | cccgcaacca | ctccagcaga | ttcttccgga | cgagatcgag | atccctaaac |
| gtgaatccgc | tcagagagaa | agagagataa | agtttaggta | acaatgtaaa | gacgagaaat |
| gtggtagaag | gtggccgcaa | gaggaagtgc | aggcattgat | ctcaagcggc | agtggcagca |
| aagagaagac | ggggatcaac | aagggagcga | tttgggatga | aagttcgaga | agcgatgtgg |
| aaagagggtg | cgaaagatct | gcgaaaaagt | gtaaggagaa | gatatcagca | agaatgaaag |
| actataggag | agtgacggaa | ggtgggcaga | aacagcctga | gtggggagaac | atgaacaagt |
| actttgagaa | acttggaat | ttttacaaga | ccatttcctc | gcacagcaag | actcgctcat |
| tgaaagattt | taaatttagg | tggttttggc | acgcaaaacg | gggagagagg | gaaaaatgag |
| cctcgagtgt | aatttttata | tcttttggtg | agttttataat | ggagaacttg | tagatgatta |
| atgtgtagaa | ggttcgaata | aaaaagacaa | atagcttggg | ttaaaactct | acggctctgt |
| gctaaggagg | acgaggagaa | ggatcctcgg | tcacatcgat | gtgattggga | ttttgtaacg |
| acttgtgagg | tctgaaatta | caaatgctga | cacttgccaa | tatggctgcc | acgttggtga |
| ttactctttc | ttctctctca | ttcattctc | ttcttcaa | cactattagc | tttattccaa |
| ggttattatt | atttataggg | atattcacia | acacaaaagt | gcttctta | ttcgggcatt |
| | | | | cgtgtattta | gaacaagaaa |

```

gatatggaac gtggaggctt ccatggctac cgcaagctgt ccgtgaacaa caccactcct
tctccaccag gtagtgccat tctctatacc cctcttttc acaggctctc ttcatttcag
ttgcatgoga aaccattctc tgcaatccct ccattgtcat gtctgtactc tttcatgac
gaacagttaa tgaaatagct tttcaatcct ataaaccgcg catgcagacg tcatcgaagc
cattatgcac taaaacttcc atttttctta tttttgttag gattagcagc gaattttctg

```

SEQ ID NO:24: 3' untranslated region

```

ga acaatggcta ataacataga cagctgacag agtcataact
gtagtaggt gcaagctgta gcttatgaat tcaagtttaa gcgaaaacaa tgctgctttt
tctttgttta ttatctatct agttgaaaga acattgtgtt tttcatctga tctgtcttgt
ggtaaagtat gtcaataaag cattagtttt gcaaaccgca tgcattgtgat attacaaaat
tcacggtgaa ttcgtaatgc gtcttggttc aaaatagaaa gagactaaac attccagatt
tcaatttctca gctacagaaa tgagtgttta acggatacag aaacaactct cacaatcttc
attcatttca tttagctact actttccaaa ggaacttcaa cgcatacctt tttcctctcc
agaagatcat gtttgtctgc actctcgttt gcctcagtat ctttctcctg atgctcttca
gatataatgtt ccaatttcga acaatcaaca ggatcaagtc cggttctttt cctctgagga
atcacagtga agaaggctgt tttccagtcc ctagtctcca gaaacttgac gagtatctcc
aaaacttggg tcacagtgag aacctaaatc aataaaaacc acaaacttta cattaacaaa
gtacataaag tagaggtttt ttgtgttgtg cccaatgaga caagaattga agtggccatt
tagttacctg agaacttgac attttcatat actctcctat gggagctta gctgttttaa
tgccttggtc ttgagccttg gtcattggtga tccctttgaa ccggtttcga tccactaagc
cacgataat gtagatatgc ttaggggtcaa gatcatccaa aacagtttca gaatcagccg
taagatacac caaattatct ttctgatcag ccatggcttc aatgtaacac ctactttcct
tttcaatgaa ccatttctca aaaccaggaa gcttgtcaag ctcagtactc atcttccc

```

WHAT IS CLAIMED IS:

1. An expression cassette comprising a promoter operably linked to a heterologous polynucleotide sequence, or a complement thereof, encoding a LEC1 polypeptide, comprising a subsequence at least 68% identical to the B domain of SEQ ID NO:2, wherein the polynucleotide sequence is heterologous to any element in the expression cassette.

2. The expression cassette of claim 1, wherein the B domain comprises a polypeptide sequence between about amino acid residue 28 and about residue 117 of SEQ ID NO:2.

3. The expression cassette of claim 1, wherein the B domain comprises a polypeptide sequence with an amino terminus at amino acid residues 28-35 and a carboxy terminus at amino acid residues 103-117 of SEQ ID NO:2.

4. The expression cassette of claim 1, wherein the LEC1 polypeptide is SEQ ID NO: 20.

5. The expression cassette of claim 4, wherein the polynucleotide sequence is SEQ ID NO:19.

6. The expression cassette of claim 1, wherein the polynucleotide sequence encodes a fusion between two or more LEC1 polypeptides or polypeptide subsequences.

7. The expression cassette of claim 1, wherein the LEC1 polypeptide is SEQ ID NO: 22.

8. The expression cassette of claim 6, wherein the polynucleotide sequence is SEQ ID NO:21.

9. The expression cassette of claim 1, wherein the promoter is a constitutive promoter.

10. The expression cassette of claim 1, wherein the promoter is from a LEC1 gene.

11. The expression cassette of claim 10, wherein the promoter comprises from about nucleotide 1 to about nucleotide 1998 of SEQ ID NO:3.

12. The expression cassette of claim 10, wherein the promoter comprises SEQ ID NO:23.

13. The expression cassette of claim 12, wherein the promoter further comprises SEQ ID NO:24.

14. The expression cassette of claim 1, wherein the polynucleotide sequence is linked to the promoter in an antisense orientation.

15. An expression cassette comprising a promoter operably linked to a heterologous polynucleotide sequence, or a complement thereof, encoding a LEC1 polypeptide comprising a subsequence at least 90% identical to the A or C domain of a LEC1 polypeptide, wherein the polynucleotide sequence is heterologous to any element in the expression cassette.

16. The expression cassette of claim 15, wherein the polynucleotide encodes a fusion between two or more LEC1 polypeptides or polypeptide subsequences.

17. An expression cassette for the expression of a heterologous polynucleotide in a plant cell, wherein the expression cassette comprises a promoter polynucleotide at least 70% identical to SEQ ID NO:23 and wherein the promoter polynucleotide is operably linked to a heterologous polynucleotide.

18. The expression cassette of claim 17, wherein the promoter comprises SEQ ID NO:23.

19. The expression cassette of claim 17, wherein the promoter further comprises a polynucleotide at least 70% identical to SEQ ID NO:24.

20. The expression cassette of claim 19, wherein the promoter comprises SEQ ID NO:24.

21. An isolated nucleic acid or complement thereof, encoding a LEC1 polypeptide comprising a subsequence at least 68% identical to the B domain of SEQ ID NO:2, with the proviso that the nucleic acid is not clone MNJ7.

22. The isolated nucleic acid of claim 21, wherein the B domain comprises a polypeptide sequence with an amino terminus at amino acids 28-35 and a carboxy terminus at amino acids 103-117 of SEQ ID NO:2.

23. The isolated nucleic acid of claim 21, wherein the LEC1 polypeptide is SEQ ID NO: 20.

24. The isolated nucleic acid of claim 23, wherein the polynucleotide sequence is SEQ ID NO:19.

25. The isolated nucleic acid of claim 21, wherein the nucleic acid encodes a fusion between two or more LEC1 polypeptides or polypeptide subsequences.

26. The isolated nucleic acid of claim 21, wherein the LEC1 polypeptide is SEQ ID NO: 22.

27. The isolated nucleic acid of claim 26, wherein the polynucleotide sequence is SEQ ID NO:21.

28. The isolated nucleic acid of claim 21, wherein the nucleic acid further comprises a promoter operably linked to the LEC1-encoding nucleic acid.

29. The isolated nucleic acid of claim 29, wherein the promoter is a constitutive promoter.

30. The isolated nucleic acid of claim 29, wherein the plant promoter is from a LEC1 gene.

31. The isolated nucleic acid of claim 30, wherein the promoter comprises from about nucleotide 1 to about nucleotide 1998 of SEQ ID NO:3.

32. The isolated nucleic acid of claim 30, wherein the promoter comprises SEQ ID NO:23.

33. The isolated nucleic acid of claim 32, wherein the promoter further comprises SEQ ID NO:24.

34. The isolated nucleic acid of claim 21, wherein the polynucleotide sequence is linked to the promoter in an antisense orientation.

35. A host cell comprising an expression cassette according to any of claims 1, 15 and 17 or a nucleic acid molecule according to claim 21, wherein the expression cassette or nucleic acid molecule is flanked by heterologous sequence.

36. The host cell of claim 35, comprising an expression cassette of claim 1.

37. The host cell of claim 35, comprising an expression cassette of claim 15.

38. The host cell of claim 35, comprising an expression cassette of claim 17.

39. The host cell of claim 35, comprising a nucleic acid molecule of claim 21.

40. An isolated polypeptide comprising an amino acid sequence

(a) at least 68% identical to the B domain of SEQ ID NO:2; and

(b) capable of exhibiting at least one of the biological activities of the polypeptide encoded by SEQ ID NO:1, SEQ ID NO: 19 or SEQ ID NO:21, or a fragment thereof.

41. An antibody capable of binding the isolated polypeptide of claim 40.

42. A method of introducing an isolated nucleic acid into a host cell comprising:

(a) providing an expression cassette according to any of claims 1, 15 and 17 or an isolated nucleic acid according to claim 21; and

(b) contacting the expression cassette or nucleic acid with the host cell under conditions that permit insertion of the nucleic acid into the host cell.

5

15

30

10

25

57. The method of claim 56, wherein the isolated nucleic acid molecule and the sample are contacted under conditions which permit the formation of a duplex between complementary nucleic acid sequences.

58. A transgenic plant cell or transgenic plant comprising the recombinant expression cassette of claim 1.

59. The transgenic plant cell or transgenic plant of claim 58, wherein the LEC1 polypeptide is SEQ ID NO:20.

60. The transgenic plant cell or transgenic plant of claim 59, wherein the polynucleotide sequence is SEQ ID NO:19.

61. The transgenic plant cell or transgenic plant of claim 58, wherein the LEC1 polypeptide is SEQ ID NO:22.

62. The transgenic plant cell or transgenic plant of claim 61, wherein the polynucleotide sequence is SEQ ID NO:21.

63. The transgenic plant cell or transgenic plant of claim 58, wherein the promoter is a constitutive promoter.

64. The transgenic plant cell or transgenic plant of claim 58, wherein the promoter comprises a promoter from a LEC1 gene.

65. The transgenic plant cell or transgenic plant of claim 58, wherein the polynucleotide sequence is linked to the promoter in an antisense orientation.

66. The transgenic plant cell or transgenic plant of claim 64, wherein the promoter comprises from about nucleotide 1 to about nucleotide 1998 of SEQ ID NO:3.

67. The transgenic plant cell or transgenic plant of claim 64, wherein the promoter comprises SEQ ID NO:23.

68. The transgenic plant cell or transgenic plant of claim 67, wherein the promoter further comprises SEQ ID NO:24.

69. A plant which has been regenerated from a plant cell according to 58.

LEAFY COTYLEDON1 GENES AND THEIR USES

ABSTRACT OF THE DISCLOSURE

5

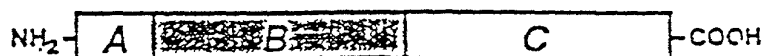
The present invention provides nucleic acid sequences from embryo-specific genes. The nucleic acids are useful in targeting gene expression to embryos or in modulating embryo development.

10

15

SF 1072651 v1

A



B

3

DNA binding

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---------------|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LEC1 | 28 | R | E | E | D | I | Q | Y | M | P | I | A | N | V | I | R | I | M | R | K | T | L | P | S | H | A | K | I | S | D | D | A |
| Maize | 30 | R | E | E | D | R | F | L | P | I | A | N | I | L | S | R | I | M | K | K | A | I | P | A | N | G | K | I | A | K | D | A |
| Chicken | 51 | R | E | E | D | I | Y | L | P | I | A | N | V | A | R | I | M | K | N | A | I | P | O | T | G | K | I | A | K | D | A | |
| Lamprey | 54 | R | E | E | D | I | Y | L | P | I | A | N | V | A | R | I | M | K | T | S | I | P | S | S | G | K | I | A | K | D | A | |
| Xenopus | 1 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | | |
| Human | 53 | R | E | E | D | I | Y | L | P | I | A | N | V | A | R | I | M | K | N | A | I | P | O | T | G | K | I | A | K | D | A | |
| Mouse/Rat | 53 | R | E | E | D | I | Y | L | P | I | A | N | V | A | R | I | M | K | N | A | I | P | O | T | G | K | I | A | K | D | A | |
| E. nidulans | 42 | K | E | E | D | R | W | L | P | I | A | N | V | A | R | I | M | K | L | A | L | P | E | N | A | K | I | A | K | E | A | |
| S. pombe | 11 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | | |
| S. cerevisiae | 36 | R | E | E | D | R | W | L | P | I | A | N | V | A | R | I | M | K | N | T | L | P | S | A | K | V | S | K | D | A | | |
| K. lactis | 21 | A | E | E | D | R | W | L | P | I | A | N | V | A | R | I | M | K | N | T | L | P | A | T | T | K | V | S | K | D | A | |

Subunit Interaction

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---------------|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LEC1 | 58 | K | E | T | I | Q | E | C | V | S | E | F | I | S | F | I | T | S | E | A | S | E | R | O | T | H | Q | E | R | K | R | K | T |
| Maize | 60 | K | E | T | V | Q | E | C | V | S | E | F | I | S | F | I | T | S | E | A | S | E | R | O | T | H | Q | E | R | K | R | K | T |
| Chicken | 81 | K | E | C | V | Q | E | C | V | S | E | F | I | S | F | I | T | S | E | A | S | E | R | O | T | H | Q | E | R | K | R | K | T |
| Lamprey | 84 | K | E | C | V | Q | E | C | V | S | E | F | I | S | F | I | T | S | E | A | S | E | R | O | T | H | Q | E | R | K | R | K | T |
| Xenopus | 1 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | | |
| Human | 83 | K | E | C | V | Q | E | C | V | S | E | F | I | S | F | I | T | S | E | A | S | E | R | O | T | H | Q | E | R | K | R | K | T |
| Mouse/Rat | 83 | K | E | C | V | Q | E | C | V | S | E | F | I | S | F | I | T | S | E | A | S | E | R | O | T | H | Q | E | R | K | R | K | T |
| E. nidulans | 72 | K | E | C | M | Q | E | C | V | S | E | F | I | S | F | I | T | S | E | A | S | E | R | O | T | H | Q | E | R | K | R | K | T |
| S. pombe | 36 | K | E | C | V | Q | E | C | V | S | E | F | I | S | F | I | T | S | E | A | S | E | R | O | T | H | Q | E | R | K | R | K | T |
| S. cerevisiae | 66 | K | E | C | M | Q | E | C | V | S | E | F | I | S | F | I | T | S | E | A | S | E | R | O | T | H | Q | E | R | K | R | K | T |
| K. lactis | 51 | K | E | C | M | Q | E | C | V | S | E | F | I | S | F | I | T | S | E | A | S | E | R | O | T | H | Q | E | R | K | R | K | T |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---------------|-----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LEC1 | 88 | I | L | A | E | O | I | L | F | A | M | S | T | L | G | F | E | N | Y | V | D | P | L | T | V | E | L | N | R | Y | R | |
| Maize | 90 | I | N | G | E | O | I | L | F | A | M | A | T | L | G | F | E | D | Y | V | E | P | L | K | V | Y | L | O | X | Y | R | |
| Chicken | 111 | I | N | G | E | O | I | L | F | A | M | S | T | L | G | F | E | D | Y | V | E | P | L | K | V | Y | L | O | X | Y | R | |
| Lamprey | 114 | I | N | G | E | O | I | L | F | A | M | S | T | L | G | F | E | D | Y | V | E | P | L | K | V | Y | L | O | X | Y | R | |
| Xenopus | 28 | I | N | G | E | O | I | L | F | A | M | S | T | L | G | F | E | D | Y | V | E | P | L | K | V | Y | L | O | X | Y | R | |
| Human | 113 | I | N | G | E | O | I | L | F | A | M | S | T | L | G | F | E | D | Y | V | E | P | L | K | V | Y | L | O | X | Y | R | |
| Mouse/Rat | 113 | I | N | G | E | O | I | L | F | A | M | S | T | L | G | F | E | D | Y | V | E | P | L | K | V | Y | L | O | X | Y | R | |
| E. nidulans | 102 | V | I | N | G | E | O | I | L | F | A | M | S | T | L | G | F | E | N | Y | A | E | V | L | K | I | S | L | T | X | Y | R |
| S. pombe | 66 | I | N | G | E | O | I | L | F | A | M | S | T | L | G | F | E | N | Y | A | E | V | L | K | I | S | L | T | X | Y | R | |
| S. cerevisiae | 98 | I | N | G | E | O | I | L | F | A | M | S | T | L | G | F | E | N | Y | A | E | V | L | K | I | S | L | T | X | Y | R | |
| K. lactis | 81 | I | N | G | E | O | I | L | F | A | M | S | T | L | G | F | E | N | Y | A | E | V | L | K | I | S | L | T | X | Y | R | |

Amino Acid Sequence Similarity between LEC1 and Other
CBF HAP3 Homologs

00515052-030400

DECLARATION

As a below named inventor, I declare that:

My residence, post office address and citizenship are as stated below next to my name; I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural inventors are named below) of the subject matter which is claimed and for which a patent is sought on the invention entitled: **LEAFY COTYLEDON1 GENES AND THEIR USES** the specification of which X is attached hereto or _____ was filed on _____ as Application No. _____ and was amended on _____ (if applicable).

I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, Section 1.56. I claim foreign priority benefits under Title 35, United States Code, Section 119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed.

Prior Foreign Application(s)

| Country | Application No. | Date of Filing | Priority Claimed Under 35 USC 119 |
|---------|-----------------|----------------|--------------------------------------|
| | | | |

I hereby claim the benefit under Title 35, United States Code § 119(e) of any United States provisional application(s) listed below:

| Application No. | Filing Date |
|-----------------|-------------|
| | |

I claim the benefit under Title 35, United States Code, Section 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, Section 112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, Section 1.56 which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

| Application No. | Date of Filing | Status |
|-----------------|----------------|---------|
| 09/193,931 | 11/17/98 | Pending |
| 09/103,478 | 6/24/98 | Pending |
| 09/026,221 | 2/19/98 | Pending |
| 08/804,534 | 2/21/97 | Pending |

| | | | | |
|-----------------------------|--|---|---|------------------------------|
| Full Name of Inventor 1: | Last Name: HARADA | First Name: JOHN | Middle Name or Initial: | |
| Residence & Citizenship: | City: Davis | State/Foreign Country: California | Country of Citizenship: United States | |
| Post Office Address: | Post Office Address: 115 Grande Avenue | City: Davis | State/Country: California | Postal Code: 95616 |

[illegible]

I further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

| | | |
|-------------------------|-------------------------|-------------------------|
| Signature of Inventor 1 | Signature of Inventor 2 | Signature of Inventor 3 |
| JOHN HARADA | TAMAR LOTAN | MASA-AKI OHTO |
| Date | Date | Date |

| | | |
|-------------------------|-------------------------|-------------------------|
| Signature of Inventor 4 | Signature of Inventor 5 | Signature of Inventor 6 |
| ROBERT B. GOLDBERG | ROBERT L. FISCHER | ANH THU BUI |
| Date | Date | Date |
| Signature of Inventor 7 | | |
| RAYMOND KWONG | | |
| Date | | |

SF 1072510 v1